

earthUI: uma interface interativa para o pacote earth (MARS)

Tradução para o português, 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: [10.67330/7fj41b65](https://doi.org/10.67330/7fj41b65)

© 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. earthUI é uma interface gráfica do usuário para o pacote **earth** do R, que implementa Multivariate Adaptive Regression Splines (MARS). 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 configuração do modelo, pelo ajuste e pela interpretação dos diagnósticos e das saídas gráficas do earth. Este artigo documenta os requisitos de formato de dados do earthUI, o fluxo de trabalho de modelagem, as telas de saída e a referência completa de funcionalidades.

Palavras-chave: MARS, splines de regressão adaptativa multivariada, earth, R, interface gráfica do usuário, avaliação de imóveis

Nota do editor (julho de 2026): Este artigo descreve o earthUI 0.10.0. Pouco antes da publicação, o earthUI 0.11.0 absorveu os aplicativos complementares glmnetUI e mgcvUI (descritos nos dois artigos seguintes), consolidando os fluxos de trabalho de MARS, elastic net e GAM em um único aplicativo. A interface e o fluxo de trabalho aqui descritos não são afetados pela consolidação.

Sumário

1	Introdução	4
1.1	O que é o earthUI?	4
1.2	Um pacote, três rotinas	4
1.3	Três modos de finalidade	4
1.4	Recursos específicos para imóveis	5
1.5	Instalação do earthUI e do vProlog	5
1.6	Primeiros passos	6
2	Requisitos dos dados de entrada do MLS	7
2.1	Formato e estrutura do arquivo	7
2.2	Colunas necessárias no modo de avaliação	7
2.3	Convenções de nomenclatura de colunas especiais	8
2.4	Qualidade e completude dos dados	8
2.5	Posicionamento da linha do avaliando	9
3	Modo de finalidade geral	10
3.1	Visão geral	10
3.2	O fluxo de trabalho da barra lateral	10
3.3	Abas do painel principal	10

3.4	Persistência das configurações	11
3.5	Modo escuro	11
3.6	Localidade e configurações regionais	11
3.6.1	Países suportados	12
3.6.2	Sobreposição do papel	12
3.6.3	Salvando padrões	12
4	Modo de avaliação	14
4.1	Tratamento da linha do avaliando	14
4.2	Data de referência e idade da venda	14
4.3	Designações especiais de colunas	14
4.4	Contribuição residual (estimação de características raras)	15
4.5	Visão geral dos ajustes RCA	16
5	Modo de análise de região de mercado	17
5.1	Diferenças em relação ao modo de avaliação	17
5.2	Quando usar o modo de mercado	17
6	Seleção de variáveis	18
6.1	Variável(is)-resposta	18
6.2	A tabela de preditores	18
6.3	Deteção de tipos de dados e sobreposições	19
6.4	Marcadores Factor e Linear	19
6.5	Referência de tipos especiais de coluna	19
6.6	Múltiplas respostas	20
7	Seleção de parâmetros	21
7.1	Parâmetros básicos	21
7.2	Passagem direta (forward pass)	21
7.3	Interações permitidas	22
7.4	Poda	22
7.5	Validação cruzada	22
7.6	Filtragem por subconjunto	23
7.7	Valores recomendados de parâmetros	23
7.7.1	Parâmetros da passagem direta	23
7.7.2	Parâmetros de poda	25
7.7.3	Parâmetros de validação cruzada	25
7.7.4	Modelo de variância	26
7.7.5	Arquivo de saída do Earth	26
7.8	Padrões das configurações	26
8	Ajustando o modelo Earth	27
8.1	O botão Fit	27
8.2	Janela modal de ajuste e saída de rastreamento	27
8.3	Log de ajuste	27
9	Baixando os dados	28
9.1	Colunas de saída	28
9.2	Ordenação das colunas e formatação do Excel	28
9.3	Escores CQA	29
9.4	Ordenação e linha do avaliando	29
9.5	Exportação automática para o mgcvUI	29

9.6	Importação no glmnetUI	30
9.6.1	Principais benefícios	30
9.6.2	Importante: a coluna de pesos precisa coincidir	30
9.6.3	Fluxo de trabalho	30
10	Cálculos RCA e download	31
10.1	Abrindo a caixa de diálogo RCA	31
10.2	Interpolação do escore CQA	31
10.3	Colunas de saída	31
10.4	A aba RCA Adjustments	32
11	Grade de comparação de vendas	33
11.1	Visão geral e finalidade	33
11.2	Janela modal de seleção de comparáveis	33
11.3	Leiaute da grade	33
11.4	Preço de venda, concessões e Net SP	34
11.5	Linhas agrupadas (Location, Site, Age)	34
11.6	Linhas das variáveis do modelo	34
11.7	Linhas CQA Residual e de características residuais	35
11.8	Fórmula do preço de venda ajustado	35
11.9	Proteção da planilha e células editáveis	35
11.10	Trabalhando com a grade no Excel	35
12	Baixando relatórios	37
12.1	Generate Quarto Report	37
12.2	Convert Quarto Report	37
12.3	Conteúdo do relatório	37
13	Os Post Processing Modules: elastic net e GAM	39
13.1	Por que ajustes de apoio?	39
13.2	Inicialização e estado compartilhado	39
13.3	A rotina de elastic net	39
13.4	A rotina de GAM	39
13.5	Um fluxo de trabalho prático	40
14	Conjunto de dados de demonstração: Appraisal_1.csv	41
14.1	Descrição	41
14.2	Colunas	42
14.3	Início rápido sugerido	42
15	Experimental: extração de texto com Prolog e regras de derivação	44
15.1	O que ele faz	44
15.2	Requisitos: vProlog (Trealla), e não SWI-Prolog	45
15.3	Habilitando e executando	45
15.4	Regras de derivação	45
15.5	Arquivo de dados ativo versionado	46
15.6	Interface programática e arquivos incluídos	46
15.7	Experimento relacionado: prioris por região climática	46
15.7.1	Além da Califórnia: pacotes de jurisdição e módulos de região de mercado	46

1 Introdução

1.1 O que é o earthUI?

earthUI é uma interface gráfica do usuário para o pacote **earth** do R, que implementa Multivariate Adaptive Regression Splines (MARS) — splines de regressão adaptativa multivariada (Friedman 1991). O motor de modelagem é o pacote **earth** de Stephen Milborrow (Milborrow 2024), documentado em detalhe por Milborrow (s.d.[a]) e Milborrow (s.d.[b]). Ele roda como um aplicativo Shiny local — não há login, nem servidor, nem contas. Você o inicia a partir do R, importa um conjunto de dados (CSV ou Excel), configura seu modelo e o ajusta interativamente.

O aplicativo oferece um fluxo de trabalho completo: importação de dados, configuração de variáveis, ajuste do modelo, gráficos de diagnóstico, importância das variáveis, equações do modelo e relatórios para download em formato HTML, Word ou PDF.

1.2 Um pacote, três rotinas

A partir da versão atual, o pacote earthUI também *contém* seus dois Post Processing Modules (módulos de pós-processamento), que em versões anteriores eram os pacotes separados glmnetUI e mgcvUI. Instalar o earthUI instala as três rotinas; cada uma roda como seu próprio aplicativo Shiny local:

- **earthUI::launch()** — o aplicativo earth (MARS) descrito neste artigo, na porta 7878. **earth()** é o motor central de avaliação.
- **earthUI::launch_glmnet()** — o Post Processing Module de elastic net (porta 7879), que importa as funções de base do modelo earth e as reestima sob regularização lasso/elastic net (veja “glmnetUI Import” adiante).
- **earthUI::launch_mgcv()** — o Post Processing Module de GAM (porta 7880), que usa as posições dos nós do modelo earth como pontos de partida para os termos suavizados do mgcv (veja “mgcvUI Auto-Export” adiante).

As três rotinas compartilham uma mesma árvore de projetos, um mesmo banco de dados de configurações e uma mesma cópia da infraestrutura de apoio, de modo que um projeto aberto em um aplicativo fica imediatamente disponível para os demais. As rotinas glmnet e mgcv são evidência de apoio à conclusão de valor do modelo earth, e não fluxos de trabalho concorrentes: o padrão previsto é ajustar e refinar primeiro o modelo earth e, depois, executar os módulos sobre o mesmo projeto quando se deseja corroboração.

1.3 Três modos de finalidade

O earthUI organiza o trabalho em **projetos** — entidades nomeadas que mantêm juntos seus arquivos de entrada, saídas e configurações em um local fixo. Todo projeto carrega uma **Purpose** (finalidade), escolhida no assistente New Project quando o projeto é criado. Há três modos:

- **General** (geral) — Regressão Earth para qualquer tipo de população ou conjunto de dados. Este é o modo padrão. Fornece o fluxo de trabalho completo de modelagem MARS, sem acréscimos específicos de domínio.
- **For Appraisal** (para avaliação) — Regressão Earth adaptada à avaliação de imóveis. Acrescenta recursos específicos da avaliação de um único imóvel, incluindo o tratamento do imóvel avaliando, as designações especiais de colunas, o Residual Constraint Approach (RCA) (abordagem por restrição residual) e a geração do Sales Comparison Grid (grade de comparação de vendas).
- **Market Area Analysis** (análise de região de mercado) — Regressão Earth adaptada a estudos de regiões de mercado. Acrescenta recursos para analisar grupos de imóveis em um mercado definido, em vez de avaliar um único imóvel avaliando.

Nos três modos, o motor central de modelagem é idêntico — você está sempre ajustando um modelo Earth (MARS). A configuração de finalidade controla quais ferramentas e elementos de interface adicionais ficam disponíveis.

A finalidade é fixa por projeto. A finalidade de um projeto — junto com seu país, estado, município e cidade — fica travada quando o projeto é criado. Mudar para um projeto diferente redefine o earthUI para um estado padrão limpo (os dados importados, os resultados do modelo, as abas, a configuração de variáveis e os parâmetros do earth são todos limpos) e então restaura a configuração salva daquele projeto para sua finalidade.

1.4 Recursos específicos para imóveis

Quando **For Appraisal** ou **Market Area Analysis** está selecionado, o earthUI ativa vários recursos concebidos para a análise de imóveis:

- **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, a saída e a geração do Sales Grid (grade de vendas). Veja o Capítulo 6 para a referência completa.
- **Arredondamento de latitude e longitude** — Colunas designadas como latitude ou longitude são automaticamente arredondadas para 3 casas decimais, a fim de evitar sobreajuste.
- **Coluna Sale Age** — Quando uma coluna é designada como `contract_date` e uma Effective Date (data de referência) é informada, o earthUI calcula uma coluna `sale_age` (dias entre a data da venda e a data de referência) e a substitui como preditor.
- **Cálculos RCA (somente avaliação)** — No modo de avaliação, depois de ajustar o modelo, o earthUI pode calcular a saída do Residual Constraint Approach (RCA), que produz ajustes por comparável, resumos de ajustes líquidos/brutos e um preço de venda ajustado para o imóvel avaliando.
- **Sales Comparison Grid (somente avaliação)** — Gera uma pasta de trabalho do Excel com várias planilhas e fórmulas para os ajustes dos comparáveis, detalhamentos residuais por característica e preços de venda ajustados. Veja o Capítulo 11.

***Dica:** O modo de finalidade 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 Instalação do earthUI e do vProlog

O earthUI está publicado no CRAN, de modo que a instalação padrão é uma linha no R:

```
install.packages("earthUI")
```

Este artigo descreve a versão atual. Se o CRAN ainda estiver distribuindo uma versão mais antiga quando você ler isto, a versão de desenvolvimento pode ser instalada a partir do GitHub (observe o `subdir` — o pacote fica na pasta `pkg/` do repositório):

```
remotes::install_github("wcraytor/earthUI", subdir = "pkg")
```

vProlog (opcional). O recurso experimental de processamento de observações (veja “Prolog Text Extraction” perto do fim deste artigo) usa o pacote vProlog — um motor Prolog embarcado e autocontido. Ele deliberadamente *não* é uma dependência obrigatória: o earthUI instala e roda sem ele, e os recursos de Prolog simplesmente permanecem desabilitados até que ele esteja

presente. O vProlog é distribuído pelo repositório r-universe do autor, que fornece binários pré-compilados para Windows e macOS (sem necessidade de cadeia de compilação):

```
install.packages("vProlog",  
                 repos = c("https://wcraytor.r-universe.dev",  
                           getOption("repos")))
```

Não é necessária nenhuma instalação de Prolog no sistema (SWI-Prolog ou outra) — o motor Trealla é compilado dentro do próprio pacote. Todo o restante deste artigo funciona sem o vProlog.

1.6 Primeiros passos

Para usar o earthUI:

1. **Instale o pacote** a partir do CRAN, conforme descrito acima.
2. **Inicie o aplicativo:** execute `earthUI::launch()` no R, ou, pela linha de comando, execute `Rscript -e 'earthUI::launch()'`. O aplicativo abre em seu navegador na porta 7878. Você também pode acessar o aplicativo diretamente navegando até `http://localhost:7878` em seu navegador. (Os Post Processing Modules iniciam da mesma forma: `launch_glmnet()` na porta 7879, `launch_mgcv()` na porta 7880.)
3. **Crie ou abra um projeto** na Seção 1 da barra lateral. O assistente New Project registra o nome do projeto, a Purpose (General, For Appraisal ou Market Area Analysis), a localização (país, estado, município, cidade) e, opcionalmente, um arquivo de dados inicial, que é copiado para a pasta `in/` do projeto.
4. **Importe seus dados** na Seção 2, que lista os arquivos CSV e Excel presentes na pasta `in/` do projeto ativo. Arquivos adicionados manualmente pelo Finder ou pelo Explorer aparecem depois de clicar em Refresh.
5. **Configure as variáveis** — escolha sua(s) variável(is)-resposta e os preditores, defina os tipos de dados e atribua quaisquer papéis especiais de coluna.
6. **Defina os parâmetros do Earth** — degree, nprune, filtros de subconjunto e outros argumentos de `earth()`.
7. **Ajuste o modelo** — clique em “Fit Earth Model” e examine os resultados no painel principal.
8. **Exporte** — baixe as previsões em Excel, gere pacotes de relatório Quarto ou (no modo de avaliação) calcule os ajustes RCA e as Sales Comparison Grids.

A configuração é salva no servidor, em um banco de dados de projeto, indexada por projeto e finalidade, por meio dos botões “Save current as default” das Seções 3 e 4 (veja o Capítulo 3, “Persistência das configurações”).

2 Requisitos dos dados de entrada do MLS

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

2.1 Formato e estrutura do arquivo

O earthUI aceita arquivos **CSV** e **Excel** (.xlsx, .xls). Na importação, os nomes das colunas são automaticamente convertidos para **snake_case** — por exemplo, “Living SqFt” torna-se **living_sqft**, “Contract Date” torna-se **contract_date** e “Sale Price” torna-se **sale_price**. Essa normalização garante referências consistentes de colunas em todo o fluxo de trabalho. O separador de CSV e o marcador decimal usados durante a 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 earthUI funcione com qualquer conjunto de colunas, o fluxo de trabalho completo de avaliação (ajustes RCA + Sales Comparison Grid) se beneficia da presença das seguintes colunas em sua exportação do MLS:

Coluna	Tipo especial	Finalidade
Sale Price	(resposta)	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; exibido no Sales Grid
Concessions	concessions	Concessões na venda; Net SP = Sale Price – Concessions no Sales Grid
Living Area (SF)	living_area	Habilita resíduos por pé quadrado (residual_sf , cqa_sf)
Lot Size	lot_size	Agrupada na linha "Site Size Dimensions" do Sales Grid
Site Dimensions	site_dimensions	Agrupada com o tamanho do lote no Sales Grid
Latitude	latitude	Arredondada para 3 casas decimais; usada no cálculo de proximidade e no grupo Location
Longitude	longitude	Arredondada para 3 casas decimais; usada no cálculo de proximidade e no grupo Location
Area ID	area	Agrupada na linha "Loc: Long Lat Area" do Sales Grid
Actual Age	actual_age	Agrupada na linha "Actual Age Effective Age" do Sales Grid
Effective Age	effective_age	Agrupada com a idade real no Sales Grid
Parcel Number	(nenhum)	Exibida na linha APN do Sales Grid
Listing ID	(nenhum)	Exibida na linha APN do Sales Grid
Address	display_only	Exibida na linha Address do Sales Grid; excluída do modelo

Os nomes de colunas da planilha podem estar em outro idioma — os nomes “especiais” estão em inglês para que o programa em R possa dar-lhes tratamento especial. No mais, os nomes de coluna informados aparecem nos modelos de regressão, nos gráficos e (no caso de avaliações) na Intermediate Sales Grid.

Nem todas as colunas são obrigatórias. O earthUI se adapta — se uma coluna estiver ausente, o recurso correspondente é simplesmente omitido. Por exemplo, se nenhuma coluna **concessions** for designada, a linha Net SP do Sales Grid mostra o Sale Price sem dedução de concessões. Contudo, para modelos de precificação imobiliária certas colunas são altamente recomendáveis para se obter um ajuste aceitável:

1. **Sale Age** — o número de dias entre a data de contrato da venda e a data de referência da avaliação ou análise. Se for usado um histórico de vendas de vários anos, especialmente para períodos superiores a 5 anos, **sale_age** costuma desempenhar papel central na estimativa do preço de venda. Na verdade, com frequência é tão importante que, sem ele, o earthUI não consegue fornecer modelo algum.
2. **Living Area** — também aparece com nomes como “Living Sqft”, “GLA” (gross living area) e assim por diante. Este é outro determinante de primeira ordem do preço de venda.
3. **Total Bath Count** — o número total de banheiros completos, de um quarto, de meio e de três quartos. Por exemplo, dois banheiros completos e um lavabo corresponderiam a um valor de 2,5.
4. **Garage Bays** ou **Garage Area** — o número de vagas de garagem ou a área da garagem.
5. **Lot Size** — a área de terreno do imóvel, tipicamente em pés quadrados ou acres.
6. **Longitude**, **Latitude** e, se disponível, **Area ID**. Os ajustes correspondentes serão combinados em um único ajuste de Location no Sales Grid.

2.3 Convenções de nomenclatura de colunas especiais

O earthUI identifica as colunas por sua **designação de tipo especial**, e não pelo nome da coluna. Você pode nomear suas colunas como quiser na exportação do MLS — o que importa é atribuir o tipo especial correto na tabela de Variable Configuration (Capítulo 6).

Por exemplo, seu MLS pode exportar a área construída como “GLA”, “Living SqFt”, “liv_area” ou “gross_living_area”. Depois da importação (quando o nome vira snake_case), basta designá-la como **living_area** no menu suspenso Special. O earthUI então a usará para cálculos de resíduo por pé quadrado e para o agrupamento no Sales Grid, independentemente de seu nome original.

2.4 Qualidade e completude dos dados

- **Valores ausentes (NA)**: linhas com valores NA em qualquer coluna preditora ou de resposta são automaticamente removidas antes do ajuste. O **na.action** é sempre definido internamente como **na.fail**, com a remoção de NA feita antes da chamada a **earth()**.
- **Colunas de data**: as datas devem estar em um formato que o R consiga interpretar (p. ex., 2025-06-15, 06/15/2025, June 15, 2025). O earthUI detecta automaticamente colunas de data quando pelo menos 80% dos valores são interpretados com sucesso. Anos com dois dígitos (p. ex., 06/15/25) são lidos como o ano correspondente mais próximo, de modo que exportações recentes do MLS são interpretadas corretamente; anos com quatro dígitos são sempre tomados literalmente.
- **Colunas numéricas**: preço de venda, área construída, tamanho do lote, concessões e campos semelhantes precisam ser numéricos. Se seu MLS exportar preços com símbolos monetários ou separadores de milhar (p. ex., “\$350,000” ou “350.000,00”), talvez seja necessário limpá-los antes da importação.
- **Colunas de fator**: variáveis categóricas como identificador de região, estilo ou estado de conservação devem conter um número gerenciável de valores distintos. O earthUI detecta

fatores automaticamente, mas você pode sobrepor a detecção.

Dica: Revise a coluna T/F/NA na tabela de preditores após a importação — ela mostra as contagens de true/false/NA de cada coluna. Colunas com mais de 30% de valores ausentes são sinalizadas em vermelho. Considere excluir colunas com muitos NA ou limpar os dados antes da importação.

2.5 Posicionamento da linha do avaliando

No modo **Appraisal**, a linha 1 deve ser o imóvel avaliando. Todas as linhas restantes são vendas comparáveis. O preço de venda do avaliando pode ser deixado em branco (NA) ou definido com qualquer valor — o earthUI o trata como NA durante o ajuste, de todo modo.

No modo **Market Area Analysis**, colocar um avaliando na linha 1 é opcional, e nenhuma linha é excluída automaticamente. Para manter uma linha de avaliando (ou qualquer outra linha) fora do ajuste, atribua a ela valor zero em uma coluna de pesos designada ou use um filtro de subconjunto.

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

3 Modo de finalidade geral

3.1 Visão geral

O modo General Purpose é o padrão quando você inicia o earthUI. Ele oferece o fluxo de trabalho completo de modelagem MARS para qualquer conjunto de dados — não apenas imóveis. Você pode usar o earthUI para dados científicos, análise financeira, estudos de engenharia ou qualquer problema de regressão em que queira explorar relações não lineares e interações entre variáveis.

No modo General, a interface omite os recursos específicos de imóveis (colunas especiais, idade da venda, arredondamento de coordenadas, RCA). A barra lateral é simplificada para focar em seleção de variáveis, configuração de parâmetros, ajuste do modelo e exportação.

3.2 O fluxo de trabalho da barra lateral

A barra lateral é organizada em seções numeradas e recolhíveis que conduzem você da seleção do projeto até a exportação:

1. Project — Escolha um projeto existente ou crie um novo pelo assistente New Project, que registra a finalidade, o país e a localização administrativa (todos travados após a criação). O projeto ativo determina de onde os arquivos de entrada são lidos e para onde todas as saídas são gravadas — não existe um campo separado de pasta de saída.

2. Import Data — Escolha um arquivo CSV ou Excel na pasta **in/** do projeto ativo; um link Refresh reexamina a pasta em busca de arquivos adicionados manualmente. Para arquivos Excel com várias planilhas, aparece um seletor de planilha. Os nomes das colunas são automaticamente convertidos para `snake_case`.

3. Variable Configuration — Seletor da variável-resposta (admite múltiplas respostas), Effective Date e a tabela de preditores com caixas de seleção Include, Factor, Linear e Latent. Veja o Capítulo 6 para os detalhes completos.

4. Earth Call Parameters — Todos os argumentos da função `earth()`: degree, penalty, nk, método de poda, validação cruzada e outros. Veja o Capítulo 7 para a referência completa dos parâmetros.

Execute Prolog Processing — Uma seção opcional e sem numeração, entre as Seções 4 e 5, desabilitada a menos que o recurso experimental de Prolog esteja habilitado em Settings. Veja o Capítulo 15.

5. Fit Earth Model — Um campo de semente aleatória e o botão que executa o modelo de forma assíncrona. Veja o Capítulo 8 para os detalhes do ajuste.

6. Download Estimated Target Variable(s) & Residuals — (intitulada “Download Estimated Sale Prices & Residuals” nos modos de avaliação e de mercado) Exporta previsões, resíduos, escores CQA e contribuições por função g como um arquivo Excel. Veja o Capítulo 9.

7–8. RCA Adjustments e Sales Grid — Somente no modo de avaliação. Veja os Capítulos 10 e 11.

9. Generate Quarto Report — Grava um pacote de relatório Quarto autocontido (fonte .qmd mais recursos) na pasta de saída do projeto. Numerada como 7 fora do modo de avaliação. Veja o Capítulo 12.

10. Convert Quarto Report — Renderiza qualquer arquivo .qmd em HTML, Word ou PDF. Veja o Capítulo 12.

3.3 Abas do painel principal

Após o ajuste, o painel principal oferece dez abas:

Aba	Conteúdo
Data	Pré-visualização dos dados importados. No modo de avaliação, dividida nas tabelas Subject Property (imóvel avaliando) e Comparable Sales (vendas comparáveis).
Equation	A equação do modelo ajustado exibida em LaTeX/MathJax, mostrando cada função de base e seu coeficiente.
Summary	Métricas principais (R^2 , GRSq, GCV, RSS, CV R^2) e uma tabela de coeficientes.
Variable Importance	Gráfico de barras e tabela ordenada dos escores de importância dos preditores.
Contribution	Gráficos de contribuição das funções g. Selecione uma função g no menu suspenso; preditores univariados exibem gráficos de linha, e interações bivariadas ($\text{degree} \geq 2$) exibem gráficos de perspectiva 3D e de contorno.
Correlation	Mapa de calor das correlações entre preditores.
Diagnostics	Três gráficos: Resíduos vs. Ajustados, Q-Q normal e Observado vs. Previsto.
RCA Adjustments	(somente avaliação) Histogramas de Residual Adj %, Net Adj % e Gross Adj % entre os comparáveis, preenchidos após a conclusão do Passo 7 (Calculate RCA Adjustments).
Residual Contribution	Estima as contribuições de mercado de características raras ou excluídas, regredindo conjuntamente os resíduos do modelo sobre colunas que o modelo nunca utilizou (veja subsection 4.4).
ANOVA	Tabela de decomposição ANOVA mostrando a contribuição de cada função de base para a RSS.
Earth Output	Saída de texto bruta de 'earth::summary()', incluindo os detalhes da etapa de poda.

3.4 Persistência das configurações

O earthUI salva sua configuração no servidor, em um banco de dados de projeto (`projects.sqlite` na raiz `regProj`), indexada por projeto e finalidade. O salvamento é **acionado por botão**: o botão “Save current as default” da Seção 3 salva a resposta, a data de referência e a configuração dos preditores, enquanto o botão da Seção 4 salva os parâmetros da chamada ao earth e a matriz Allowed Interactions. Cada botão atualiza apenas a sua metade, de modo que nunca se sobrepõem, e **o ajuste não salva automaticamente** — você pode experimentar livremente sem sobrescrever seus padrões salvos. Os vários arquivos de entrada de um projeto (um pequeno extrato de teste, o conjunto de dados completo) compartilham uma configuração e, como as configurações vivem na árvore do projeto, elas viajam com ela quando você sincroniza, faz backup ou compartilha a pasta `regProj`.

3.5 Modo escuro

Clique no ícone de lua/sol no canto superior direito para alternar entre os temas claro e escuro. A preferência de tema é salva no armazenamento local e persiste entre sessões.

3.6 Localidade e configurações regionais

O earthUI oferece suporte a convenções internacionais de números, datas e formatação de CSV por meio de um sistema de localidades baseado em país. O menu suspenso **Country** no menu **Settings** (engrenagem) da barra de menus superior seleciona uma predefinição entre 31 países suportados. Cada predefinição configura:

- **Separador de CSV** — vírgula (,) para EUA/Reino Unido/Japão ou ponto e vírgula (;) para a maior parte da Europa, onde a vírgula é usada como marcador decimal.
- **Marcador decimal** — ponto (.) ou vírgula (,).
- **Separador de milhar** — vírgula (EUA/Reino Unido/Japão), ponto (Alemanha/Itália/Espanha), espaço (Finlândia/França/Polônia/Bálticos/Ucrânia/Rússia) ou apóstrofo (Suíça).
- **Formato de data** — MM/DD/AAAA (EUA), DD/MM/AAAA (maior parte da Europa) ou AAAA-MM-DD (Suécia/Lituânia/Japão/Canadá).
- **Tamanho do papel** — Letter (EUA/Canadá/México) ou A4 (todos os demais).

3.6.1 Países suportados

País	Sep. CSV	Decimal	Milhar	Data	Papel	
EUA, Canadá (csv)	,	.	,	MDY/YMD	Letter	
Reino Unido, Irlanda, AU, NZ	,	.	,	DMY	A4	
Alemanha, Áustria, Bélgica, Países Baixos, Itália, Espanha, Portugal, Dinamarca, Brasil	;	,	.	DMY	A4	
França, Polônia, Rep. Tcheca, Noruega, Letônia, Estônia	;	,	(espaço)	DMY	A4	
Finlândia, Suécia	;	,	(espaço)	DMY/YMD	A4	
Suíça	;	,	'	DMY	A4	
Ucrânia, Rússia	;	,	(espaço)	DMY	A4	
Turquia	;	,	.	DMY	A4	
Japão, Coreia do Sul	,	.	,	YMD	A4	
México	,	.	,	DMY	Letter	
Lituânia	;	,	(espaço)	YMD	A4	

3.6.2 Sobreposição do papel

Abaixo do seletor de país, um menu suspenso **Paper** (Letter ou A4) permite sobrepor o tamanho de página do relatório sem trocar de país. O separador de CSV, o marcador decimal e a ordem do formato de data seguem a predefinição do país selecionado; quando você muda de país, a sobreposição de papel volta ao padrão daquele país.

3.6.3 Salvando padrões

Clique em **Save** no menu Settings para armazenar globalmente suas preferências de localidade. Esses padrões se aplicam a todas as sessões futuras, independentemente do projeto que você abrir. As configurações por projeto (resposta, preditores, parâmetros) são salvas separadamente no banco de dados do projeto; os padrões de localidade persistem em todos os projetos por meio de um banco de dados SQLite no nível do usuário. Essa abordagem em dois níveis foi pensada para organizações como empresas de auditoria que trabalham com dados de vários países — defina o país mais frequente como padrão e ajuste quando necessário.

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

4 Modo de avaliação

Quando você seleciona **For Appraisal** como Purpose, o earthUI 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 linhas restantes são vendas comparáveis. Seu arquivo de entrada deve estar organizado de acordo (veja o Capítulo 2). O preço de venda do avaliando pode ser deixado em branco ou definido com qualquer valor — o earthUI automaticamente o trata como NA durante o ajuste.

Após a importação, a aba Data se divide em duas seções: **Subject Property** (linha 1) e **Comparable Sales** (linhas 2 em diante). A linha 1 é sempre excluída do ajuste do modelo — a notificação “Skipping row 1 (subject). Fitting on N rows.” confirma isso. Após o ajuste, o modelo ainda gera previsões para a linha do avaliando, exibidas como `est_<target>` na saída.

4.2 Data de referência e idade da venda

Um campo **Effective Date** aparece na seção Variable Configuration em todos os modos de finalidade (com a data de hoje como padrão) — análises gerais e de mercado também têm uma data de referência, não apenas as avaliações — mas ele só é *usado* nos modos de avaliação e de mercado. Nesses modos, se você designar uma coluna como `contract_date` no menu suspenso Special, o earthUI calcula uma coluna `sale_age` — o número de dias inteiros entre a data de contrato de cada venda e a data de referência. Essa coluna substitui a coluna de data original como preditor.

Na primeira vez em que você clicar em Fit depois de designar uma data de contrato, o earthUI cria a coluna `sale_age` e o avisa para clicar em Fit novamente a fim de incluí-la.

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. A lista completa de tipos especiais e seus efeitos:

Tipo especial	Efeito
no	Padrão — sem tratamento especial
contract_date	Dispara o cálculo automático de <code>sale_age</code> a partir da Effective Date
listing_date	Usada com a data do contrato para calcular o DOM quando não existe coluna <code>dom</code>
dom	Identifica a coluna de dias no mercado; exibida no Sales Grid
concessions	Identifica as concessões da venda; usada na fórmula Net SP (Sale Price – Concessions) do Sales Grid
latitude	Valores arredondados para 3 casas decimais; usados na proximidade de Haversine e no grupo Location do Sales Grid
longitude	Valores arredondados para 3 casas decimais; usados na proximidade de Haversine e no grupo Location do Sales Grid
area	Agrupada na linha "Loc: Long Lat Area" do Sales Grid
living_area	Habilita resíduos por pé quadrado (<code>residual_sf</code> , <code>cqa_sf</code>) na saída de download
lot_size	Agrupada na linha "Site Size Dimensions" do Sales Grid
site_dimensions	Agrupada com o tamanho do lote no Sales Grid
actual_age	Agrupada na linha "Actual Age Effective Age" do Sales Grid
effective_age	Agrupada com a idade real no Sales Grid
display_only	Coluna incluída nas exportações, mas excluída do ajuste do modelo (permite várias)
remarks	(Experimental) Coluna de texto livre de observações, convertida em colunas de características pela etapa opcional de Prolog (permite várias); veja o Capítulo 15
list	(Experimental) Coluna delimitada por vírgulas, dividida em colunas TRUE/FALSE one-hot pela etapa opcional de Prolog (permite várias); veja o Capítulo 15

Só é permitida uma coluna por tipo especial (exceto `display_only`, `remarks` e `list`). Atribuir um especial a uma segunda coluna automaticamente o remove da primeira. Um pequeno selo azul aparece ao lado do nome da variável, mostrando o tipo especial atribuído.

4.4 Contribuição residual (estimação de características raras)

Características raras demais para o canal da regressão — uma oficina presente em um punhado de vendas — pertencem à camada residual do RCA. A aba **Residual Contribution** estima sua contribuição de mercado regredindo os resíduos do modelo ajustado sobre colunas que o modelo nunca utilizou, por mínimos quadrados ordinários. Selecionar várias colunas as estima *conjuntamente*, de modo que amenidades correlacionadas (p. ex. `outbuilding_sqft` mais `pr_has_workshop`) dividem o crédito em vez de serem contadas em dobro. Indicadores binários produzem estimativas de valor global e listam o resíduo de cada venda sinalizada como evidência; colunas numéricas produzem contribuições por unidade; o erro padrão reportado é um EP ($t \geq 2$ como critério usual de significância). Termos apoiados em menos de dez observações recebem automaticamente uma RESSALVA na nota gerada para o dossiê, que documenta método, estimativas e advertências em um único parágrafo para o relatório. Somente os preditores que o modelo efetivamente *usou* são excluídos da lista de candidatos — um indicador raro oferecido ao ajuste, mas nunca selecionado pela passagem direta, permanece elegível, pois o modelo não absorveu nada de seu valor.

4.5 Visão geral dos ajustes RCA

O botão **Calculate RCA Adjustments & Download** (seção 7 da barra lateral, visível apenas no modo de avaliação após o ajuste) calcula ajustes derivados do mercado para cada comparável em relação ao avaliando. O fluxo de trabalho completo do RCA é descrito no Capítulo 10.

Depois de calcular os ajustes RCA, o botão **Generate Sales Grid & Download** (seção 8 da barra lateral) fica disponível. O fluxo de trabalho do Sales Grid é descrito no Capítulo 11.

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

5 Modo de análise de região de mercado

Quando você seleciona **Market Area Analysis** como Purpose, o earthUI oferece os mesmos recursos específicos de imóveis do modo de avaliação (colunas especiais, idade da venda, arredondamento de coordenadas), mas orientado à análise de um grupo de imóveis, em vez da avaliação de um único avaliando.

5.1 Diferenças em relação ao modo de avaliação

- **Sem exclusão automática do avaliando** — diferentemente do modo de avaliação, o modo de mercado não exclui automaticamente a linha 1 do ajuste. Todas as linhas são incluídas; para manter uma linha de avaliando fora do ajuste, atribua a ela valor zero em uma coluna de pesos designada (como no conjunto de dados de demonstração) ou use um filtro de subconjunto.
- **Sem seções de RCA ou Sales Grid** — as etapas “Calculate RCA Adjustments & Download” e “Generate Sales Grid & Download” não estão disponíveis. O modo de mercado concentra-se no ajuste do modelo e na saída, e não em ajustes por comparável.
- **Numeração da barra lateral** — sem as etapas de RCA e Sales Grid, a seção Generate Quarto Report é numerada como 7 em vez de 9 (a seção Convert Quarto Report continua sendo a 10).

5.2 Quando usar o modo de mercado

O modo Market Area Analysis é apropriado quando você está:

- Construindo um modelo de regressão para um bairro ou região de mercado a fim de compreender os fatores determinantes de valor
- Analisando como variáveis como área construída, idade, tamanho do lote e localização afetam os preços de venda de um grupo de imóveis
- Preparando fundamentação para uma análise de condições de mercado ou uma delimitação de bairro
- Trabalhando com um conjunto de dados que inclui um imóvel avaliando na linha 1, mas querendo a opção de incluí-lo ou excluí-lo do ajuste

Dica: O modo de mercado também é útil para regressões imobiliárias em geral, quando você quer os recursos de colunas especiais (idade da venda, arredondamento de coordenadas), mas não precisa do fluxo de trabalho de ajustes RCA.

6 Seleção de variáveis

A Seção 3 da barra lateral — **Variable Configuration** — é onde você escolhe quais colunas participam do modelo e como elas são tratadas.

6.1 Variável(is)-resposta

O menu suspenso **Target (response) variable(s)**, no topo da Seção 3, lista todas as colunas do seu conjunto de dados. Selecione uma coluna para um modelo padrão de resposta única, ou várias colunas para um modelo de múltiplas respostas. Quando vários alvos são selecionados, o modelo ajusta todas as respostas simultaneamente usando `earth(cbind(y1, y2, ...) ~ .)`.

As colunas selecionadas como resposta são automaticamente excluídas da lista de preditores.

6.2 A tabela de preditores

Abaixo do seletor de resposta, uma tabela lista todas as colunas restantes com os seguintes campos:

Coluna	Descrição
Variable	Nome da coluna (nome completo exibido na dica de tela se truncado), precedido de um alternador $\times$ / $+$ (veja abaixo). Nos modos de avaliação/mercado, um selo azul mostra o tipo especial atribuído.
Type	Menu suspenso de tipo de dados: <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>logical</code> , <code>factor</code> , <code>Date</code> , <code>POSIXct</code>
Include	Caixa de seleção — incluir esta coluna como preditor no modelo
Factor	Caixa de seleção — tratar esta coluna como variável categórica
Linear	Caixa de seleção — forçar entrada apenas linear (sem funções dobradiça)
Latent	Caixa de seleção — manter esta coluna inteiramente fora do modelo (excluída do <code>earth</code> e dos dados exportados para o <code>glmnetUI</code> / <code>mgcvUI</code>), reservando-a para trabalho com variáveis latentes
Special	Menu suspenso (somente avaliação/mercado) — veja a Referência de tipos especiais de coluna, abaixo
T/F/NA	Três contagens por coluna: para lógicas, TRUE/FALSE/NA; para numéricas, $\neq 0$ / $= 0$ /NA. Exibidas em vermelho quando mais de 30% dos valores estão ausentes

Uma linha de dica acima da tabela explica as abreviações: “Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. `contract_date`).”

Desabilitação suave em vez de exclusão. O alternador $\times$ ao lado de cada nome de variável esmaece a linha e mantém a coluna fora do modelo sem removê-la dos dados; ele se transforma em um $+$ verde que reabilita a coluna. Colunas nunca são destruídas.

6.3 Detecção de tipos de dados e sobreposições

O earthUI detecta automaticamente os tipos de dados na importação. Colunas numéricas, inteiras, lógicas, de fator e de data são reconhecidas. Colunas de caracteres que parecem datas (pelo menos 80% dos valores interpretáveis segundo formatos de data comuns) são classificadas como **Date**.

Você pode sobrepor qualquer detecção alterando o menu suspenso **Type**. Ao mudar uma coluna para **character** ou **factor**, a caixa Factor é marcada automaticamente. Alterar tipos afeta como a coluna é passada à função **earth()**.

6.4 Marcadores Factor e Linear

- **Factor** — Quando marcado, a coluna é tratada como variável categórica. Isso é apropriado para colunas como **style**, **area_id** ou **grade**, que representam grupos discretos, e não medições contínuas. Colunas de fator entram no modelo como variáveis indicadoras (*dummy*).
- **Linear** — Quando marcado, a coluna é forçada a entrar no modelo linearmente — nenhuma função dobradiça (linear por partes) é criada para ela. Isso é útil quando você sabe que uma variável tem relação estritamente linear com a resposta.

6.5 Referência de tipos especiais de coluna

Nos modos de avaliação e de mercado, o menu suspenso **Special** oferece as seguintes opções. Cada tipo pode ser atribuído a no máximo uma coluna (exceto **display_only**, **remarks** e **list**, que permitem várias):

Tipos de data e hora:

- **contract_date** — Dispara o cálculo automático de **sale_age**. A coluna de data original é substituída por uma coluna inteira que mede os dias entre a data da venda e a Effective Date.
- **listing_date** — Usada como alternativa para calcular os dias no mercado (DOM = data do contrato – data do anúncio) quando nenhuma coluna **dom** explícita é designada.
- **dom** — Identifica a coluna de dias no mercado. Exibida na linha APN e na linha Date of Sale do Sales Grid.

Tipos monetários:

- **concessions** — Identifica concessões da venda (créditos do vendedor, incentivos ao comprador etc.). Usada no Sales Grid para calcular o preço de venda líquido: Net SP = Sale Price – Concessions.

Tipos de tamanho e localização:

- **latitude** — Os valores são automaticamente arredondados para 3 casas decimais, a fim de evitar sobreajuste. Usados nos cálculos de proximidade de Haversine (distância do avaliando a cada comparável) e agrupados na linha Location do Sales Grid.
- **longitude** — Mesmo tratamento de arredondamento da latitude. Usada para proximidade e para o grupo Location.
- **area** — Tipicamente um identificador de bairro ou região. Agrupada com latitude e longitude na linha “Loc: Long | Lat | Area” do Sales Grid.
- **living_area** — Habilita os cálculos de resíduo por pé quadrado (**residual_sf** e **cqa_sf**) na saída de download.
- **lot_size** — Agrupada na linha “Site Size | Dimensions” do Sales Grid.
- **site_dimensions** — Agrupada com o tamanho do lote no Sales Grid (p. ex., “75x120”).

Tipos de idade:

- **actual_age** — Agrupada na linha “Actual Age | Effective Age” do Sales Grid.
- **effective_age** — Agrupada com a idade real no Sales Grid.

Tipos de exibição:

- **display_only** — A coluna é incluída nas exportações para Excel, mas totalmente excluída do ajuste do modelo. Use isto para campos de endereço, números do MLS, identificadores de matrícula ou outros dados de referência que não devem ser preditores. Várias colunas podem receber essa designação.

Tipos de extração de texto (experimentais):

- **remarks** — Uma coluna de texto livre (p. ex. **public_remarks**) a ser convertida em colunas de características pela etapa opcional de Prolog. Não faz nada até que esse recurso seja habilitado em Settings. Várias colunas são permitidas. Veja o Capítulo 15.
- **list** — Uma coluna delimitada por vírgulas (p. ex. **existing_structures**) a ser dividida em colunas TRUE/FALSE <col>_<value> one-hot pela mesma etapa. Várias colunas são permitidas. Veja o Capítulo 15.

6.6 Múltiplas respostas

Selecionar mais de uma variável-resposta ajusta um modelo Earth de múltiplas respostas. Quando vários alvos são selecionados:

- O modelo prevê todas as respostas simultaneamente, compartilhando funções de base entre os alvos
- O botão **wp (response weights)** torna-se ativo, permitindo que você atribua um peso numérico a cada alvo
- Os modelos de variância (**varmod.method**) são desabilitados (sem suporte para múltiplas respostas)
- As abas de resultados exibem métricas, equações e gráficos de diagnóstico por resposta

***Dica:** Colunas designadas como **display_only** permanecem no seu conjunto de dados e aparecem na exportação para Excel, mas são inteiramente excluídas da tabela de preditores. Use isto para colunas de identificação, endereços ou outros campos de referência.*

7 Seleção de parâmetros

A Seção 4 da barra lateral — **Earth Call Parameters** — dá acesso a todos os argumentos aceitos pela função `earth()`. Cada parâmetro tem um ícone azul de ajuda (?) com uma explicação em dica de tela. Os parâmetros são organizados em subseções recolhíveis.

7.1 Parâmetros básicos

Parâmetro	Padrão	Descrição
subset	(vazio)	Expressão de filtro de linhas. Veja “Filtragem por subconjunto”, abaixo.
weights	NULL	Seletor de coluna para pesos de caso (linha). Somente colunas numéricas são listadas.
wp	NULL	Pesos das respostas em modelos de múltiplos alvos. O botão abre uma caixa de diálogo com uma entrada numérica por alvo (padrão 1.0 cada).
keepxy	desligado	Reter x, y, subset e weights no objeto do modelo.
trace	0	Nível de rastreamento da saída de ajuste (0 a 5).
glm	nenhum	Família GLM opcional: gaussian , binomial ou poisson .
degree	1	Ordem máxima de interação. Definir $\text{degree} \geq 2$ habilita automaticamente a validação cruzada e revela a matriz Allowed Interactions.
penalty	2	Penalidade GCV por nó. Valores mais altos produzem modelos mais simples.

7.2 Passagem direta (forward pass)

Parâmetro	Padrão	Descrição
nk	automático	Máximo de termos antes da poda. Preenchido automaticamente com o valor recomendado para os dados carregados (veja “Valores recomendados de parâmetros”, abaixo).
thresh	0.001	Limiar da passagem direta. Valores menores permitem mais termos.

Parâmetro	Padrão	Descrição
minspan	automático	Vão mínimo entre nós. Preenchido automaticamente com o valor recomendado. Valores negativos definem o número máximo de nós por preditor.
endspan	automático	Vão de extremidade — distância mínima de um nó à borda dos dados. Preenchido automaticamente com o valor recomendado.
newvar.penalty	0	Penalidade por introduzir uma nova variável (incentiva o reaproveitamento de preditores já existentes).
fast.k	20	Número de termos-pai considerados no algoritmo fast MARS.
fast.beta	1	Controla o fator de envelhecimento do fast MARS.

7.3 Interações permitidas

Quando **degree** é definido como 2 ou mais, uma matriz de interações aparece abaixo dos parâmetros básicos. Trata-se de uma grade triangular superior de caixas de seleção, uma para cada par de preditores. Uma caixa marcada significa que os dois preditores podem interagir; desmarcá-la proíbe aquela interação específica.

Clicar no nome de um preditor (rótulo de linha ou coluna) alterna todas as interações daquela variável. As caixas **Allow All** e **Clear All** no topo fornecem controle em massa. A matriz usa cabeçalhos fixos, de modo que os rótulos de linha e coluna permanecem visíveis durante a rolagem.

Um alerta informativo lembra: “Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold).” (Os termos de interação aumentam o risco de sobreajuste. A validação cruzada foi habilitada, com 10 dobras.)

7.4 Poda

Parâmetro	Padrão	Descrição
pmethod	backward	Método de poda: backward , none , exhaustive , forward , seqrep ou cv .
nprune	NULL	Máximo de termos após a poda. Deixe vazio para seleção automática.

7.5 Validação cruzada

Parâmetro	Padrão	Descrição
nfold	10	Número de dobras da validação cruzada. Defina 0 para desabilitar a VC.
ncross	20	Número de repetições da validação cruzada.
stratify	ligado	Estratifica as amostras da VC para que cada dobra tenha distribuição de resposta semelhante.

Quando $\text{degree} \geq 2$, a validação cruzada é habilitada automaticamente (nfold definido como 10) para ajudar a proteger contra o sobreajuste decorrente das interações.

7.6 Filtragem por subconjunto

A entrada de texto **subset** aceita uma expressão R que filtra quais linhas são usadas no ajuste do modelo. Você pode digitar uma expressão diretamente (p. ex., `sale_age < 365 & area_id == 460`) ou usar o botão **Build filter...** para construí-la visualmente.

Caixa de diálogo Build Filter — Clique em “Build filter...” para abrir uma caixa de diálogo guiada. Cada linha de condição tem um menu suspenso de coluna, um operador (<, >, <=, >=, ==, !=) e uma entrada de valor que se adapta ao tipo da coluna: entrada numérica para números, seletor de data para datas e menu suspenso de valores distintos para colunas de caractere/fator. As condições são unidas por conectores AND (&) ou OR (|). Uma pré-visualização na parte inferior mostra a expressão e quantas linhas correspondem. Clique em **Apply** para inserir a expressão na entrada de texto.

Colunas de data precisam usar os invólucros `as.Date("...")` ou `as.POSIXct("...")` em expressões manuais. A caixa de diálogo Build Filter cuida disso automaticamente.

A filtragem por subconjunto não é destrutiva — as linhas excluídas permanecem no conjunto de dados e recebem previsões na exportação para Excel.

***Dica:** Use parênteses para agrupar condições OR ao combiná-las com AND. Sem parênteses, $A \& B \mid C$ é avaliado como $(A \& B) \mid C$, o que pode não ser o pretendido.*

7.7 Valores recomendados de parâmetros

O earthUI exibe um valor recomendado abaixo dos parâmetros principais da Seção 4. Essas recomendações são atualizadas reativamente com base no número de linhas usadas no ajuste (n) e de preditores selecionados (p). As fórmulas derivam do artigo de Friedman sobre MARS, dos algoritmos internos do earth e de testes empíricos. Os campos **nk**, **minspan** e **endspan** são pré-preenchidos com seus valores recomendados à medida que os dados são carregados; uma edição manual é preservada.

7.7.1 Parâmetros da passagem direta

nk (máximo de termos antes da poda):

$$\text{nk} = \min(100, \max(21, 2p + 1, \lfloor n/10 \rfloor))$$

n	$p=5$	$p=10$
30	21	21
50	21	21
100	21	21
200	21	21
500	50	50
1000	100	100
1500	100	100

O padrão do earth, $\min(200, \max(20, 2p)) + 1$, não leva em conta o tamanho do conjunto de dados e pode ser pequeno demais para conjuntos grandes, restringindo a passagem direta antes que ela explore adequadamente todos os preditores. O termo $\lfloor n/10 \rfloor$ permite aproximadamente um termo a cada 10 observações — uma regra prática padrão para evitar superparametrização. O teto de 100 evita retornos decrescentes.

minspan (mínimo de observações entre nós):

$$\text{minspan} = \min(16, \lfloor 5 + n/50 \rfloor)$$

endspan (mínimo de observações a partir das bordas dos dados):

$$\text{endspan} = \min(16, \lfloor 5 + n/28 \rfloor)$$

n	minspan	endspan
30	5	6
50	6	6
100	7	8
200	9	12
300	11	15
500	15	16
1500	16	16

O minspan autocalculado pelo earth usa a equação 43 de Friedman: $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$, que cresce como $\ln(np)$ e aumenta lentamente demais para conjuntos de dados grandes. Em $n=1500$ ela produz apenas ≈ 7 , o que dá cerca de 245 posições candidatas a nó por preditor contínuo — muitas, permitindo que a passagem direta ajuste ruído. A fórmula recomendada visa aproximadamente 100 candidatos por preditor para n grande, deixando a cargo do autocálculo do earth os casos de n pequeno (nos quais ele é bem testado). O endspan é definido um pouco maior que o minspan para oferecer proteção adicional nas bordas, o que ajuda quando os dados têm caudas ralas (comum em imóveis).

penalty (penalidade GCV por nó):

$$\text{penalty} = \begin{cases} 3 & \text{se degree} > 1 \\ 2 & \text{se degree} = 1 \end{cases}$$

Este é o próprio padrão do earth — 2 para modelos aditivos, 3 para modelos com interações (a penalidade mais alta compensa o espaço de busca maior).

newvar.penalty (penalidade por introduzir um novo preditor):

Recomendado: **0,1** quando os preditores são correlacionados (p. ex., área construída com número de dormitórios, número de banheiros, tamanho do lote). Isso envia a passagem direta

em direção ao reaproveitamento de preditores já presentes no modelo, em vez de introduzir alternativas correlacionadas. O mecanismo: a melhora de RSS de cada novo preditor é multiplicada por $1/(1 + \text{penalty})$. Com `newvar.penalty = 0.1`, uma variável nova precisa melhorar a RSS em pelo menos 10% mais do que outro nó em uma variável já existente. Isso produz modelos mais simples, com menos preditores, sem afetar as estimativas finais dos coeficientes (a penalidade é removida após a seleção). Deixe em 0 se os preditores não forem correlacionados. Note que o `earth` não admite um `newvar.penalty` diferente de zero junto com pesos de caso — o `earthUI` desabilita a entrada (forçando 0) sempre que uma coluna de pesos é designada.

7.7.2 Parâmetros de poda

pmethod: Recomendado: **backward**. A poda regressiva baseada em GCV, que é o padrão, é determinística — os mesmos dados sempre produzem o mesmo modelo, sem dependência de sementes aleatórias. Isso é importante para a reprodutibilidade, especialmente em trabalhos de avaliação. O método `cv` usa validação cruzada para a poda, o que introduz dependência da semente.

nprune: Recomendado: **deixar vazio** (NULL). Deixe o GCV selecionar o número ótimo de termos. Definir `nprune` impõe um teto rígido que se sobrepõe ao critério do GCV.

7.7.3 Parâmetros de validação cruzada

nfold (dobras da VC): recomendado por faixa de tamanho amostral:

n (linhas do ajuste)	nfold
0–299	5
300–399	6
400–499	7
500–599	8
600–699	9
700+	10

Mais dobras treinam cada dobra com mais dados (estimativa menos enviesada), mas encolhem a dobra de teste retida. A tabela evolui até as 10 dobras padrão à medida que a amostra cresce, mantendo cada dobra de teste grande o bastante (cerca de 50+ linhas) para uma métrica por dobra estável, e recua para 5 dobras em amostras pequenas — nas quais a estimativa de CVR^2 ainda é ruidosa, de modo que convém aumentar `ncross` (p. ex. 15–20) para diluir a variância. O teto de 10 reflete o fato de que cada dobra adicional é um ajuste completo do `earth` com ganho diagnóstico desprezível. Com `pmethod = "backward"`, a validação cruzada não afeta o modelo — ela apenas calcula o CVR^2 diagnóstico e fornece resíduos para o modelo de variância.

ncross (repetições da VC): recomendado por faixa de tamanho amostral:

n (linhas do ajuste)	ncross
0–199	20
200–399	16
400–699	12
700–1199	10
1200–2999	7
3000+	5

As repetições diluem a sorte do particionamento em dobras e estabilizam o CVR^2 ; amostras maiores precisam de menos repetições porque uma única validação cruzada já é mais estável. Os

resíduos da VC repetida também alimentam o modelo de variância.

7.7.4 Modelo de variância

varmod.method: Recomendado: **lm** — o padrão de Milborrow: um modelo linear robusto de variância para intervalos de previsão (a dispersão é estimada a partir de resíduos ruidosos, de modo que um modelo simples generaliza melhor). Se **lm** não convergir com seus dados, experimente **rlm** ou **const** (homocedástico, sempre converge), ou recorra a **earth** (que também lida com dispersão não linear). Ele exige **nfold** > 1 e **ncross** > 1.

7.7.5 Arquivo de saída do Earth

Após cada ajuste bem-sucedido, o earthUI grava <filename>_earth_output_<timestamp>.txt na pasta de saída. Esse arquivo contém os termos do modelo, estatísticas de resumo (R^2 , GRSq, CVRSq), detalhes do modelo de variância e o log de rastreamento. Um arquivo é criado por ajuste, fornecendo um registro cumulativo para comparar configurações de parâmetros.

7.8 Padrões das configurações

Um botão de opção no topo da Seção 4 controla quais padrões são carregados:

- **Use last settings for input file** — restaura as configurações que você usou da última vez com esse arquivo específico
- **Use default settings** — aplica seus padrões personalizados salvos
- **Earth defaults** — redefine todos os parâmetros para os valores de fábrica

Clique em **Save current as default** na parte inferior da Seção 4 para armazenar os parâmetros atuais do earth e a matriz Allowed Interactions; o botão correspondente na Seção 3 armazena a resposta, a data de referência e a configuração dos preditores. Ambos são salvos por projeto e finalidade no banco de dados do projeto.

8 Ajustando o modelo Earth

8.1 O botão Fit

A Seção 5 da barra lateral contém o botão **Fit Earth Model** junto com um campo **Random seed**. A semente torna as atribuições de dobras da validação cruzada reprodutíveis — um menu suspenso recupera as últimas 10 sementes usadas, e a semente utilizada é registrada na aba Earth Output após o ajuste. Antes do ajuste, o earthUI valida sua configuração — é preciso selecionar uma resposta e incluir ao menos um preditor. Nos modos de avaliação/mercado, as colunas de latitude e longitude são arredondadas, e a idade da venda é calculada a partir da data de referência e da data de contrato (se designada).

8.2 Janela modal de ajuste e saída de rastreamento

Quando você clica em Fit, aparece uma sobreposição modal escura com:

- **Cabeçalho** — “Fitting Earth Model”, com um cronômetro de tempo decorrido contando os segundos e um botão **Abort** que cancela o ajuste em segundo plano
- **Log de rastreamento** — uma área de texto rolável, em fonte monoespaçada, exibindo em tempo real a saída da função `earth()`: informações do conjunto de dados, progresso da passagem direta, dobras da validação cruzada e tempo de conclusão
- **Código de cores** — linhas padrão em verde, linhas de validação cruzada em amarelo, erros em vermelho

O earthUI ajusta modelos de forma assíncrona usando `callr::r_bg()`, que executa o cálculo em um processo R separado. Isso mantém o aplicativo responsivo durante ajustes demorados. Um observador com sondagem de 300 ms lê a saída do processo em segundo plano e a transmite para o log de rastreamento.

Quando o ajuste termina, aparece a mensagem “Done in X.Xs”, um botão de fechar (X) é adicionado à janela modal e o fundo escurecido é removido, de modo que as abas de resultado possam ser renderizadas sob demanda. A janela modal então se fecha automaticamente assim que as saídas visíveis terminam de ser renderizadas — ou você pode fechá-la antes.

Em caso de sucesso, uma marca de verificação verde é acrescentada ao botão Fit. Se ocorrer um erro, a mensagem de erro é exibida no log de rastreamento e aparece uma notificação.

8.3 Log de ajuste

Todo ajuste (bem-sucedido ou não) grava um arquivo de log na sua pasta de saída:

- **Nome do arquivo:** `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **Conteúdo:** carimbo de data/hora, toda a saída de rastreamento do processo de ajuste e quaisquer mensagens de erro
- **Local:** a pasta de saída do projeto ativo

***Dica:** Se o `callr` não estiver instalado, o earthUI recorre ao ajuste síncrono com uma barra de progresso simples. Instale o `callr` para ter a experiência completa de saída de rastreamento.*

9 Baixando os dados

Após o ajuste, a seção 6 da barra lateral — intitulada “Download Estimated Target Variable(s) & Residuals” no modo geral e “Download Estimated Sale Prices & Residuals” nos modos de avaliação e de mercado — baixa um arquivo Excel com previsões e diagnósticos. Essa saída é usada no Passo 7 (RCA) para atribuir uma classificação CQA (Condition-Quality-Appeal — estado, qualidade e atratividade) ao imóvel avaliando. A saída é ordenada por **residual_sf** e **cqa_sf** para ajudá-lo a avaliar em que ponto do ordenamento o avaliando se situa. Se o modelo for de boa qualidade, os imóveis devem ficar ordenados do menos atraente ao mais atraente, com base em características residuais que não entraram na regressão. O valor central deve ser aproximadamente 0, a metade inferior com valores negativos e a metade superior com valores positivos. Você deve encontrar as casas de pior qualidade, ou “para reformar”, perto do fim do ordenamento, e as melhores casas no topo. Costuma haver exceções decorrentes de anomalias como execuções hipotecárias, vendas por valor inferior à dívida, vendas de inventário (ligadas a herança) e vendas rápidas motivadas por mudança de emprego ou outras razões. A investigação das anomalias em geral revela uma razão pertinente para a anomalia de preço.

O rótulo do botão se adapta ao modo de finalidade:

- Geral/Mercado: **Download Output (Excel)**
- Avaliação: **Download Intermediate Output (Excel)**

O formato do nome de arquivo é `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`.

9.1 Colunas de saída

Para cada variável-resposta, são acrescentadas as seguintes colunas:

Coluna	Descrição
<code>est_<target></code>	Previsão do modelo, p. ex. <code>est_sale_price</code> (1 casa decimal)
<code>residual</code>	Observado menos previsto (1 casa decimal)
<code>cqa</code>	Escore Condition-Quality-Appeal (2 casas decimais, faixa 0–10)
<code>residual_sf</code>	Resíduo dividido pela área construída (se designada, 1 casa decimal)
<code>cqa_sf</code>	CQA calculado a partir do ordenamento por <code>residual_sf</code> (2 casas decimais)
<code><variable>_contribution</code>	Valor de contribuição por função g (1 casa decimal, uma coluna por função g)
<code>basis</code>	Contribuição do valor do intercepto, igual para todos os imóveis (1 casa decimal)
<code>calc_residual</code>	Observado menos (basis + todas as contribuições) — coluna de verificação (1 casa decimal)

Em modelos de múltiplas respostas, um sufixo `<i>` é acrescentado para distinguir as colunas de cada alvo.

9.2 Ordenação das colunas e formatação do Excel

As colunas de ordenamento são colocadas na **posição mais à esquerda** do arquivo de saída, nesta ordem: **residual_sf**, **cqa_sf**, **residual**, **cqa**. Isso facilita percorrer a lista ordenada e avaliar onde o avaliando se situa na distribuição de CQA.

Essas colunas recebem formatação numérica do Excel:

Coluna	Formato	Exemplo
<code>residual_sf</code>	Numérico, 2 casas decimais	12.34
<code>cqa_sf</code>	Numérico, 2 casas decimais	7.25
<code>residual</code>	Numérico, 0 casas decimais	15,200
<code>cqa</code>	Numérico, 2 casas decimais	6.80

9.3 Escores CQA

O escore CQA (Condition-Quality-Appeal) ordena o resíduo de cada linha em relação a todos os demais resíduos. Para uma dada linha, o CQA é a porcentagem de linhas com resíduo com sinal menor, multiplicada por 10. Isso produz uma escala de 0 a 10 em que:

- **CQA alto** ($\approx 9-10$) — o imóvel foi vendido por muito mais do que o modelo previa (resíduo positivo grande)
- **CQA baixo** ($\approx 0-1$) — o imóvel foi vendido por muito menos do que o previsto (resíduo negativo grande)
- **CQA ≈ 5** — o resíduo está próximo da mediana de todos os resíduos

Quando uma coluna `living_area` é designada, `cqa_sf` fornece o mesmo ordenamento com base nos resíduos por pé quadrado.

9.4 Ordenação e linha do avaliando

Nos modos de avaliação e de mercado, as linhas de comparáveis são ordenadas por `residual_sf` de forma decrescente (ou por `residual`, se nenhuma área construída for designada). A linha do avaliando (linha 1) permanece na posição 1 e não é ordenada. No modo geral, as linhas são exportadas em sua ordem original.

Essa ordenação, combinada com as colunas de ordenamento mais à esquerda, permite ao avaliador percorrer rapidamente os comparáveis, do mais superestimado ao mais subestimado, e avaliar onde o escore CQA atribuído ao avaliando se situa nessa distribuição.

Os níveis de fator nos dados de previsão são alinhados com os dados de treinamento antes da chamada a `predict()`. Linhas com níveis de fator não vistos produzirão previsões NA.

***Dica:** A coluna `calc_residual` deve sempre ser igual a `residual` (a menos de arredondamento). Se houver diferença significativa, isso indica um problema de cálculo que merece investigação.*

9.5 Exportação automática para o mgcvUI

A cada ajuste bem-sucedido com `degree` ≤ 2 , o earthUI salva automaticamente o objeto de resultado completo como um arquivo `.rds` na pasta de saída do projeto. O nome do arquivo segue o padrão `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`.

Esse arquivo pode ser carregado pelo Post Processing Module de mgcv — incluído no pacote earthUI e iniciado com `earthUI::launch_mgcv()` — usando `readRDS()`. A rotina GAM usa as posições dos nós e as funções de base do modelo earth como pontos de partida para os termos suavizados do GAM, permitindo uma transição fluida de MARS para GAM.

Modelos com `degree` > 2 são ignorados porque o mgcvUI só admite interações de duas variáveis. Quando o earthUI é iniciado no modo Trilogy (a partir da interface Trilogy), uma seção adicional **Lock Model Output** permite travar o `.rds` de um determinado ajuste — junto com sua base de preditores e as configurações de CQA compartilhadas — como o modelo earth canônico que o glmnetUI e o mgcvUI importam.

9.6 Importação no glmnetUI

O mesmo arquivo `.rds` salvo para a rotina GAM também pode ser importado no **Post Processing Module de elastic net** — igualmente incluído no pacote `earthUI` e iniciado com `earthUI::launch_glmnet()`. No aplicativo `glmnet`, navegue até **Section 2: Import from earthUI** e use o botão **Browse** para selecionar o arquivo `.rds`.

O `glmnetUI` usa a abordagem padrão `earth::model.matrix()`: as funções de base do modelo `earth` (dobradiças, interações e termos lineares) tornam-se as colunas da matriz de delineamento do `glmnet`. O `glmnet` então realiza regressão regularizada sobre essas colunas de base, selecionando-as e encolhendo-as via `lasso/elastic net`. Isso combina a construção adaptativa de base do `earth` com a regularização do `glmnet`.

9.6.1 Principais benefícios

- **Interações interpretáveis:** as interações do `earth` baseadas em dobradiças (inclusive termos de três vias) têm significado geométrico claro, ao contrário das interações nativas por produto cruzado do `glmnet`, difíceis de explicar em juízo ou em contextos de auditoria.
- **Não linearidade automática:** as funções dobradiça do `earth` capturam relações não lineares que um modelo `glmnet` padrão com preditores brutos deixaria escapar.
- **Seleção de variáveis:** o `glmnet` pode podar ainda mais a base do `earth`, descartando termos dobradiça que não contribuem.

9.6.2 Importante: a coluna de pesos precisa coincidir

Aviso

Se o `earthUI` usar uma coluna de pesos (p. ex., para excluir a linha do avaliando com `weight=0`), a mesma coluna de pesos precisa ser designada na configuração de variáveis do `glmnetUI`. Se as configurações de peso diferirem, as contagens de linhas não coincidirão e a base do `earth` não poderá ser anexada à matriz do modelo do `glmnet`.

9.6.3 Fluxo de trabalho

1. Ajuste um modelo `earth` no `earthUI` (qualquer degree).
2. O arquivo de resultado `.rds` é salvo automaticamente na pasta de saída do projeto.
3. No `glmnetUI`, abra **Section 2: Import from earthUI** e navegue até o arquivo `.rds`.
4. Verifique se o **mesmo arquivo de dados** e a **mesma coluna de pesos** são usados nos dois aplicativos.
5. Configure os parâmetros do `glmnet` na Seção 5 (a Interaction Matrix é desabilitada automaticamente quando há uma importação do `earthUI` ativa).
6. Clique em **Fit Model**. As colunas de base do `earth` aparecem nas abas **Equation**, **Coefficients** e **Contributions**.

10 Cálculos RCA e download

O botão **Calculate RCA Adjustments & Download** aparece na seção 7 da barra lateral, visível apenas no modo de avaliação depois que um modelo tenha sido ajustado. O RCA (Residual Constraint Approach) produz ajustes por comparável derivados do mercado, em relação ao imóvel avaliando.

10.1 Abrindo a caixa de diálogo RCA

Clicar no botão abre uma pequena caixa de diálogo modal com:

- **Score type** — botões de opção para escolher **CQA** ou **CQA per SF** (a opção por pé quadrado só está disponível quando uma coluna `living_area` está designada). Se você escolher “CQA per SF”, então, com base no escore CQA que atribuir ao avaliando, seu escore residual será sua área construída multiplicada pelo `residual_sf` correspondente ao escore CQA_SF informado.
- **CQA value** — entrada para o escore CQA do avaliando (0,00–10,00). Esse escore representa onde você acredita que o avaliando se situa na distribuição de qualidade dos comparáveis.
- Botão **Generate** — calcula a saída do RCA e a baixa como Excel. O botão permanece desabilitado até que um escore válido seja informado.

10.2 Interpolação do escore CQA

Quando você clica em Generate, o earthUI interpola o resíduo do avaliando a partir dos pares CQA/resíduo dos comparáveis:

1. Os escores CQA e os resíduos dos comparáveis são ordenados por CQA crescente
2. A interpolação linear (`stats::approx()`) mapeia o valor de CQA que você informou para um resíduo
3. Se estiver usando CQA por pé quadrado, o resíduo por pé quadrado é convertido de volta em resíduo total pela multiplicação pela área construída do avaliando
4. O valor estimado do avaliando é calculado como: previsão do modelo + resíduo interpolado

10.3 Colunas de saída

O arquivo Excel do RCA (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) inclui todas as colunas da saída intermediária mais:

Coluna	Descrição
<code>subject_value</code>	Previsão do modelo + resíduo interpolado (somente linha 1)
<code>subject_cqa</code>	O escore CQA que você informou (somente linha 1)
<code><variable>_adjustment</code>	Contribuição do avaliando menos contribuição do comparável (por função g)
<code>residual_adjustment</code>	Resíduo do avaliando menos resíduo do comparável
<code>net_adjustments</code>	Soma de todos os ajustes (contribuição + resíduo)
<code>gross_adjustments</code>	Soma dos valores absolutos de todos os ajustes
<code>residual_pct</code>	Ajuste residual como fração do preço de venda do comparável
<code>net_adj_pct</code>	Ajustes líquidos como fração do preço de venda do comparável

Coluna	Descrição
gross_adj_pct	Ajustes brutos como fração do preço de venda do comparável
adjusted_sale_price	Preço de venda do comparável + ajustes líquidos

As colunas de ajuste informam, para cada comparável, quanto da diferença o modelo atribui a cada variável. **adjusted_sale_price** é o preço de venda do comparável depois de aplicados todos os ajustes derivados do modelo — um conjunto de preços de venda ajustados que deve se agrupar em torno do valor estimado do avaliando.

10.4 A aba RCA Adjustments

Uma vez calculada a saída do RCA, a aba **RCA Adjustments** do painel principal exibe histogramas das distribuições de Residual Adj %, Net Adj % e Gross Adj % entre os comparáveis (um conjunto por alvo, em modelos de múltiplas respostas), oferecendo uma verificação visual rápida de quão concentrados estão os ajustes.

Dica: Um escore CQA de 5,00 posiciona o avaliando na mediana dos comparáveis. Escores acima de 5 indicam que o avaliando tem qualidade acima da média; abaixo de 5, qualidade abaixo da média.

11 Grade de comparação de vendas

11.1 Visão geral e finalidade

A **Sales Comparison Grid** (grade de comparação de vendas) é uma pasta de trabalho do Excel gerada no modo de avaliação (seção 8 da barra lateral) depois do cálculo dos ajustes RCA. Ela apresenta o imóvel avaliando ao lado dos comparáveis selecionados em um formato de grade estruturado, com fórmulas do Excel que calculam automaticamente os ajustes e um preço de venda ajustado para cada comparável.

A grade foi concebida para o dossiê do avaliador — ela combina os ajustes derivados da regressão do modelo Earth com células editáveis nas quais o avaliador pode alocar o resíduo CQA a características específicas do imóvel. O nome do arquivo de saída é `SalesGrid_<YYYYMMDD_HHMMSS>.xlsx`.

11.2 Janela modal de seleção de comparáveis

Clicar em “Generate Sales Grid & Download” abre uma caixa de diálogo modal na qual você seleciona quais comparáveis incluir:

- Os **comparáveis recomendados** já vêm marcados. São comparáveis com percentual de ajuste bruto inferior a 25% do preço de venda, ordenados por idade da venda crescente (vendas mais recentes primeiro) e limitados a 30.
- Os **comparáveis adicionais** são listados abaixo, desmarcados por padrão. Eles têm ajustes brutos de 25% ou mais, mas continuam disponíveis para seleção (até 50 exibidos, ordenados por percentual de ajuste bruto crescente).
- Podem ser selecionados no máximo **30 comparáveis**, produzindo até **10 planilhas** (3 comparáveis por planilha).

Depois de confirmar sua seleção, os comparáveis escolhidos são ordenados por percentual de ajuste bruto crescente, de modo que os comparáveis mais semelhantes apareçam na Planilha 1.

11.3 Leiaute da grade

Cada planilha tem um leiaute de 20 colunas, acomodando o imóvel avaliando mais 3 comparáveis. As colunas de cada entidade são:

Entidade	Colunas (5 por entidade)
Avaliando	Rótulo, Valor factual 1, Valor factual 2, Valor factual 3, Contribuição de valor (VC)
Cada comparável	Valor factual 1, Valor factual 2, Valor factual 3, Contribuição de valor (VC), Ajuste

As linhas, de cima para baixo, são:

1. **Título** — nome da planilha (p. ex., “Intermediate Sales Comparable Grid — Sheet 1 of 3”)
2. **Cabeçalhos** — “Subject” e os cabeçalhos das colunas dos comparáveis, com endereço
3. **Address** — endereço completo do imóvel
4. **APN | MLS# | DOM | Subj.Prox** — número de matrícula, identificador do anúncio, dias no mercado e distância de Haversine (milhas) até o avaliando
5. **Sales Price | Concess. | Net SP** — preço de venda, concessões e fórmula do preço de venda líquido
6. Linha de cabeçalho **Regression Features**
7. **BASE VALUE** — o intercepto do modelo

8. **Date of Sale | OffMkt | OnMkt** — data do contrato, idade da venda e DOM
9. **Linhas agrupadas** (condicionais) — grupos Location, Site Size e/ou Age
10. **Linhas das variáveis do modelo** — uma linha por preditor (excluindo as variáveis agrupadas)
11. **Linha separadora em branco**
12. Linha de cabeçalho **Residual Features**
13. **CQA|Residual** — escore CQA + fórmula do resíduo remanescente
14. **Linhas de características residuais** — linhas nomeadas + linhas em branco para preenchimento pelo avaliador
15. **Total VC / Net Adjustment**
16. **Net Adjustment %**
17. **Gross Adjustment %**
18. **Adjusted Sale Price** — linha de fórmula
19. Rodapé de **copyright**

11.4 Preço de venda, concessões e Net SP

A linha 5 mostra três valores para cada comparável:

- **Sale Price** — o preço de venda do comparável, conforme os dados
- **Concessions** — o valor da coluna designada como **concessions** (0 se não designada)
- **Net SP** — uma fórmula do Excel: Sale Price – Concessions

A coluna do avaliando mostra “N/A” para o preço de venda (uma vez que ele é desconhecido) e qualquer valor de concessões, se disponível.

11.5 Linhas agrupadas (Location, Site, Age)

Quando certos tipos especiais são designados e essas variáveis aparecem no modelo, o earthUI cria linhas agrupadas que combinam variáveis relacionadas:

Loc: Long | Lat | Area — Aparece quando qualquer uma de **longitude**, **latitude** ou **area** está no modelo. Mostra os valores factuais de cada variável constituinte, uma contribuição de valor combinada (soma das VCs individuais) e, para os comparáveis, um ajuste (VC combinada do avaliando – VC combinada do comparável). Estilizada com fundo azul-claro nas células de VC.

Site Size | Dimensions — Aparece quando **lot_size** ou **site_dimensions** estão no modelo. Mesma estrutura do grupo Location.

Actual Age | Effective Age — Aparece quando **actual_age** ou **effective_age** estão no modelo. Mesma estrutura.

As variáveis consumidas pelas linhas agrupadas são **excluídas das linhas individuais de variáveis do modelo** abaixo, evitando dupla contagem nos totais de ajuste.

11.6 Linhas das variáveis do modelo

Abaixo das linhas agrupadas (ou diretamente abaixo de BASE VALUE, se não houver linhas agrupadas), uma linha por preditor restante do modelo mostra:

- **Valores factuais** do avaliando e de cada comparável (os valores efetivos dos dados)
- **Contribuição de valor (VC)** — a contribuição da função g para o valor previsto daquela linha
- **Ajuste** (somente comparáveis) — VC do avaliando menos VC do comparável

11.7 Linhas CQA|Residual e de características residuais

A linha **CQA|Residual** mostra o escore CQA de cada imóvel e contém uma fórmula para o **resíduo remanescente** — a parcela do resíduo ainda não alocada a características específicas. A fórmula é:

$$\text{Resíduo remanescente} = \text{Resíduo total} - \sum (\text{VCs das características residuais abaixo})$$

Abaixo disso vêm as **linhas de características residuais** — 5 linhas nomeadas (Location, View/Appeal, Condition, Quality, Other) mais 6 linhas em branco. São células de entrada nas quais o avaliador informa as contribuições de valor de características não captadas pelo modelo. À medida que o avaliador preenche valores, a fórmula do Resíduo remanescente diminui automaticamente.

Para cada comparável, a coluna de ajuste contém uma fórmula: VC da característica do avaliando – VC da característica do comparável.

11.8 Fórmula do preço de venda ajustado

A linha **Adjusted Sale Price** contém fórmulas do Excel:

- **Avaliando:** soma de todas as células de contribuição de valor, de BASE VALUE até a última linha de característica residual. Isso representa a previsão total do modelo para o avaliando mais quaisquer características residuais alocadas pelo avaliador.
- **Comparáveis:** Net SP + soma de todas as células de ajuste, da primeira linha agrupada ou de variável até a última linha de característica residual. Isso dá o preço de venda do comparável após todos os ajustes derivados da regressão e informados pelo avaliador.

11.9 Proteção da planilha e células editáveis

Cada planilha é **protegida** para evitar a modificação acidental de fórmulas e dados. A proteção bloqueia:

- Todas as células de fórmula (Net SP, resíduo remanescente, ajustes, preço de venda ajustado etc.)
- Todas as células de dados (endereços, preços de venda, valores factuais, contribuições de valor do modelo)
- Rótulos e cabeçalhos

As únicas células **desbloqueadas** (editáveis) são as entradas de contribuição de valor das características residuais — as células sob a linha CQA|Residual nas quais o avaliador informa os detalhes. Elas são estilizadas com fundo amarelo-claro para indicar que são editáveis.

11.10 Trabalhando com a grade no Excel

1. **Abra o arquivo** no Excel ou em um aplicativo de planilha compatível.
2. **Revise os ajustes da regressão** — os ajustes derivados do modelo já vêm preenchidos e bloqueados.
3. **Aloque o resíduo** — nas linhas de características residuais (células amarelas), informe sua avaliação de quanto do resíduo remanescente é atribuível a cada característica (Location, View, Condition, Quality etc.). A fórmula do Resíduo remanescente diminuirá conforme você aloca.
4. **Verifique o Adjusted Sale Price** — essa fórmula é atualizada automaticamente conforme você informa valores de características residuais.
5. **Várias planilhas** — se você selecionou mais de 3 comparáveis, navegue entre as planilhas. Cada planilha tem o mesmo avaliando na coluna da esquerda, com 3 comparáveis diferentes.

Dica: O objetivo é alocar o resíduo CQA entre as linhas de características até que o Resíduo remanescente fique próximo de zero. Isso demonstra que você consegue explicar a diferença entre a estimativa de regressão de cada comparável e seu preço de venda efetivo.

12 Baixando relatórios

Depois de ajustar um modelo, a geração de relatórios é um fluxo de trabalho Quarto em duas etapas: **Generate Quarto Report** (seção 9 da barra lateral no modo de avaliação, 7 nos demais) grava um pacote de relatório autocontido, e **Convert Quarto Report** (seção 10) o renderiza — ou qualquer outro arquivo Quarto — nos formatos finais.

12.1 Generate Quarto Report

Clicar em **Generate Quarto Report** grava um pacote Quarto autocontido na pasta de saída do projeto, em `<base>_qmd/`, contendo a fonte `.qmd` preenchida, todos os recursos de gráfico pré-gerados (PNG e PDF), os dados do relatório e um modelo de referência do Word. Nenhuma renderização ocorre nesta etapa — o pacote pode ser editado à mão, combinado com outras fontes Quarto por meio de `includes` do Quarto, ou pré-visualizado isoladamente com `quarto preview`.

12.2 Convert Quarto Report

A Seção 10 oferece um seletor de arquivos (com o pacote gerado mais recentemente no projeto atual como padrão — mas você pode navegar até qualquer arquivo `.qmd`) e caixas de seleção de formato:

- **HTML** — um arquivo `.html` com renderização matemática KaTeX, imagens embutidas e sumário. Usa o tema Bootstrap Flatly e a fonte Roboto Condensed.
- **Word** — um arquivo `.docx` adequado para edição e distribuição. Inclui sumário, números de página e usa um modelo de referência personalizado para estilo consistente.
- **PDF** — composto com LuaLaTeX para saída de qualidade profissional. O tamanho do papel segue a configuração de localidade (Letter ou A4). Inclui páginas em paisagem para matrizes de interação grandes e usa Roboto Condensed com Latin Modern Math para as equações. A opção de PDF só é oferecida quando uma distribuição LaTeX é detectada.

Clique em **Convert** para renderizar os formatos selecionados. A renderização roda em segundo plano — uma caixa de diálogo modal mostra o tempo decorrido e o progresso do Quarto, enquanto o aplicativo permanece responsivo. Uma notificação confirma a localização do arquivo ao término.

Todas as operações (ajuste do modelo, downloads de dados, cálculos RCA, geração da grade de vendas e renderização de relatórios) são registradas com carimbos de início/fim e tempos decorridos em `<datafile>_earthui_log.txt`, na pasta de saída, para solução de problemas e monitoramento de desempenho.

12.3 Conteúdo do relatório

Todo relatório gerado inclui as seguintes seções:

1. **Descrição do conjunto de dados** — número de observações, variável(is)-resposta, preditores utilizados, preditores categóricos
2. **Especificação do modelo** — degree, situação da validação cruzada, número de termos, preditores no modelo
3. **Allowed Interactions** (somente $\text{degree} \geq 2$) — a matriz de interações com marcas de verificação, formatada conforme o tipo de saída
4. **Resumo dos resultados** — R^2 , GRSq , GCV , RSS e $\text{CV } R^2$ (se a validação cruzada estiver habilitada)
5. **Equação do modelo** — a equação completa do modelo Earth em notação LaTeX

6. **Coeficientes e funções de base** — tabela de todos os termos, seus coeficientes e as funções de base que os definem
7. **Importância das variáveis** — gráfico de barras e tabela ordenada dos escores de importância dos preditores
8. **Contribuições das funções g** — gráficos para cada função g: gráficos de linha para termos univariados, gráficos de perspectiva 3D e de contorno para interações bivariadas
9. **Matriz de correlação** — mapa de calor das correlações entre preditores
10. **Diagnósticos** — gráficos de Resíduos vs. Ajustados, Q-Q normal e Observado vs. Previsto
11. **Decomposição ANOVA** — tabela mostrando a contribuição de cada função de base para a RSS
12. **Earth Output** — saída de texto bruta de `earth::summary()`, incluindo os detalhes da etapa de poda

Em modelos de múltiplas respostas, as seções 4–8 e 10 se repetem para cada variável-resposta.

***Dica:** Os relatórios exigem o pacote R `quarto` e uma instalação do Quarto. Para saída em PDF, também é preciso ter uma distribuição LaTeX instalada (p. ex., via `tinytex::install_tinytex()`) — caso contrário, a caixa de seleção de PDF fica oculta.*

13 Os Post Processing Modules: elastic net e GAM

O pacote earthUI traz dois Post Processing Modules — uma rotina de elastic net construída sobre `glmnet` e uma rotina de GAM construída sobre `mgcv`. Cada uma é tratada em profundidade por um artigo próprio nesta edição; esta seção explica *para que* elas servem no fluxo de trabalho do earthUI e como as três rotinas cooperam.

13.1 Por que ajustes de apoio?

O modelo earth produz a conclusão de valor. Os Post Processing Modules existem para **corroborá-la**: duas estimativas adicionais do mesmo avaliando, produzidas por motores metodologicamente distintos, ajustadas aos mesmos dados do projeto. Quando uma reestimação por elastic net da base do earth e um suavizador GAM independente chegam a valores materialmente iguais, a conclusão do avaliador deixa de repousar sobre as premissas de um único modelo — uma vantagem significativa em contextos de revisão e de litígio. Quando eles *discordam*, essa discordância é ela própria diagnóstica, apontando em geral para dados escassos, um termo excessivamente flexível ou um comparável atípico que merece reexame. Nenhum dos módulos se sobrepõe ao modelo earth; eles são evidência de apoio, deliberadamente apresentada sem ponderação.

13.2 Inicialização e estado compartilhado

Cada rotina é um aplicativo Shiny local separado:

- `earthUI::launch_glmnet()` — o aplicativo de elastic net, na porta 7879.
- `earthUI::launch_mgcv()` — o aplicativo de GAM, na porta 7880.

Os três aplicativos leem a mesma árvore de projetos e o mesmo banco de dados de configurações, de modo que o projeto aberto no earthUI fica imediatamente disponível nos módulos — junto com os campos herdados que o earthUI salva para ele: a Response (resposta), a Effective Date, o tipo e o valor do CQA do avaliando no RCA, e a designação de área construída. Você não os informa de novo; os módulos os leem, para que os mesmos valores percorram os três ajustes.

13.3 A rotina de elastic net

A rotina `glmnet` importa as funções de base do modelo earth — suas dobradiças, interações e termos lineares — como colunas de uma matriz de delineamento e as reestima sob regularização lasso/elastic net (veja “`glmnetUI Import`”, acima, para a mecânica e a ressalva quanto à coluna de pesos). O resultado preserva os termos interpretáveis e geometricamente significativos do earth, deixando que o regularizador encolha ou descarte aqueles que não se justificam. Veja o artigo sobre `glmnetUI` nesta edição para o tratamento completo.

13.4 A rotina de GAM

A rotina `mgcv` segue o caminho inverso: em vez de reaproveitar a base do earth, ela ajusta novos termos suavizados penalizados, usando as posições dos nós do modelo earth como pontos de partida (veja “`mgcvUI Auto-Export`”, acima). Como o GAM rederiva as formas funcionais sob um paradigma de suavização diferente, sua concordância com o modelo earth é corroboração genuinamente independente, e não um eco. São suportados modelos earth com $\text{degree} \leq 2$. Veja o artigo sobre `mgcvUI` nesta edição para o tratamento completo.

13.5 Um fluxo de trabalho prático

1. Ajuste e refine **primeiro o modelo earth**, no earthUI, até ficar satisfeito com ele. Todo ajuste bem-sucedido ($\text{degree} \leq 2$) exporta automaticamente o **.rds** que os módulos consomem.
2. Abra as rotinas **glmnet** e **mgcv** sobre o mesmo projeto, importe o resultado do earth e ajuste. Trabalhe em uma rotina de cada vez — os ajustes são demorados e as rotinas se coordenam por meio do projeto compartilhado, de modo que o ritmo sensato é sequencial, não paralelo.
3. Itere livremente. Cada aplicativo mantém seus resultados enquanto você alterna entre eles, e nada obriga a gerar relatórios pelo caminho.
4. Somente quando os três resultados o satisfizerem, volte a cada rotina e gere seu relatório Quarto. Cada método registra seu ajuste e sua conclusão de valor no registro trilogy do projeto, mantendo os três relatórios rastreáveis a uma única execução comparativa.

Dica: Os módulos são opcionais em todos os níveis: o fluxo de trabalho próprio do earthUI — ajuste, RCA, Sales Grids, relatórios — é completo sem eles. Recorra a eles quando uma conclusão exigir corroboração: um avaliando difícil, um segmento de mercado com poucos dados ou um trabalho com probabilidade de passar por revisão.

14 Conjunto de dados de demonstração: Appraisal_1.csv

O earthUI inclui um conjunto de dados MLS de demonstração para explorar o fluxo de trabalho de avaliação. Carregue-o programaticamente com:

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

Ou, no aplicativo Shiny, copie-o para a pasta `in/` de um projeto (ou anexe-o como arquivo de dados inicial na criação do projeto) e selecione-o na Seção 2 (Import Data).

14.1 Descrição

O arquivo contém 1.501 vendas residenciais mais 1 imóvel avaliando na linha 1 (1.502 linhas ao todo), de uma exportação simulada de MLS. Os dados representam vendas de casas unifamiliares em um mercado com várias regiões, com uma amplitude de tamanhos, idades e localizações de imóveis.

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

14.2 Colunas

Coluna	Tipo	Tipo especial	Descrição
weight	numeric	—	Peso da observação (0 = excluir do ajuste)
id	numeric	display_only	Identificador interno do registro
property_id	numeric	display_only	Identificador do imóvel no MLS
listing_id	character	display_only	Número do anúncio no MLS
parcel_number	character	display_only	Número de matrícula do cadastro municipal (APN)
street_address	character	display_only	Endereço do imóvel
city_name	character	display_only	Cidade
postal_code	character	display_only	CEP
county_name	character	display_only	Município
contract_date	Date	contract_date	Data do contrato de venda (calcula sale_age)
sale_age	numeric	—	Dias da data do contrato até a data de referência (pré-calculado)
coe_date	Date	display_only	Data de fechamento da escritura
listing_status	character	display_only	Situação do anúncio (p. ex., "Sold")
sale_price	numeric	(resposta)	Preço de venda — variável-resposta
rent	numeric	—	Aluguel mensal (para modelos de múltiplas respostas)
list_price	numeric	display_only	Preço de anúncio
original_list_price	numeric	display_only	Preço de anúncio original
living_sqft	numeric	living_area	Área construída bruta em pés quadrados
beds_total	integer	—	Número de dormitórios
baths_total	numeric	—	Contagem total de banheiros (p. ex., 2.5 = 2 completos + 1 lavabo)
lot_size	numeric	lot_size	Tamanho do lote em pés quadrados
area_id	integer	area	Identificador de região do MLS
area_text	character	display_only	Nome da região
age	numeric	actual_age	Idade do imóvel em anos
year_built	integer	display_only	Ano de construção
latitude	numeric	latitude	Latitude (arredondada para 3 casas decimais no modelo)
longitude	numeric	longitude	Longitude (arredondada para 3 casas decimais no modelo)
latitude6	numeric	display_only	Latitude com precisão total
longitude6	numeric	display_only	Longitude com precisão total
garage_spaces	integer	—	Número de vagas de garagem
fp_count	integer	—	Número de lareiras
no_of_stories	numeric	—	Número de pavimentos
style	character	—	Estilo arquitetônico
view	character	—	Tipo de vista (p. ex., "Neighborhood", "Hills")
days_on_market	integer	dom	Dias no mercado
listing_date	Date	listing_date	Data do anúncio
sale_concessions	numeric	concessions	Concessões do vendedor

14.3 Início rápido sugerido

1. Inicie o earthUI: `earthUI::launch()`
2. Crie um projeto com Purpose **For Appraisal** no assistente New Project, anexando `Appraisal_1.csv` como arquivo de dados inicial
3. Selecione `Appraisal_1.csv` na Seção 2 (Import Data)
4. Selecione `sale_price` como resposta
5. Atribua os tipos especiais conforme a tabela acima
6. Inclua os preditores: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (como fator), `age`, `latitude`, `longitude`, `garage_spaces`
7. Defina degree como 1 e clique em **Fit Earth Model**
8. Baixe a saída intermediária (Passo 6) e revise o ordenamento por CQA
9. Calcule os ajustes RCA (Passo 7) com um escore CQA de ~5,00
10. Gere a Sales Comparison Grid (Passo 8)

Dica: O conjunto de dados de demonstração é uma boa forma de explorar todos os recursos do earthUI antes de importar seus próprios dados de MLS. Os preditores sugeridos acima costumam produzir um modelo com R^2 acima de 0,85.

15 Experimental: extração de texto com Prolog e regras de derivação

A versão 0.10.0 introduz um recurso **opcional e genuinamente experimental** para extrair características prontas para modelagem a partir das colunas de texto descritivo de uma exportação de MLS — sobretudo a coluna “Public Remarks”, mas também as observações do corretor e listas de características delimitadas por vírgulas, como características dos cômodos ou estruturas existentes. O motor de extração é Prolog e, em particular, **Definite Clause Grammars (DCGs)**: as regras que reconhecem expressões como “move in ready”, “3 car garage” ou “owned solar” no texto das observações são escritas como regras gramaticais Prolog.

Este recurso ainda está em desenvolvimento. Ele já pode ser usado *parcialmente* para extração hoje — a gramática incluída já reconhece um conjunto útil de características — mas adaptá-lo ao vocabulário do seu próprio MLS exige ler e escrever Prolog, e DCGs em particular. Se esse não for o seu caso, pule este capítulo por completo: o recurso vem desligado por padrão, e o restante do earthUI funciona exatamente igual sem ele.

Aviso

Este é um recurso experimental em desenvolvimento ativo. Sua gramática de extração não é completa nem validada para trabalho de avaliação em produção — as colunas extraídas precisam ser conferidas contra o texto de origem antes de serem usadas em um modelo. Usar o recurso de forma eficaz exige conhecimento prático de Prolog e DCGs. Tudo o que ele faz é aditivo e reversível; se algo se comportar mal, desligue o seletor em Settings e siga normalmente.

15.1 O que ele faz

Os anúncios de MLS carregam uma grande quantidade de informação relevante para o valor em texto livre e em campos de lista delimitados. Esta etapa opcional transforma esse texto em colunas comuns que sua regressão pode usar, por meio de dois papéis especiais de coluna (Capítulo 6):

- **Colunas remarks** são interpretadas pela gramática DCG de Prolog em uma coluna de característica por atributo extraído. O prefixo da coluna é construído a partir das primeiras letras do nome da coluna de origem: `public_remarks` gera colunas `pr_*`, `agent_remark` gera `ar_*`, e assim por diante.
- **Colunas list** (p. ex. `existing_structures = "Shed, Workshop"`) são divididas em colunas `TRUE/FALSE <col>_<value>` one-hot. Essa parte é R puro e não precisa de Prolog algum.

A gramática incluída (`inst/prolog/mls_remarks.pl`) atualmente extrai, entre outras coisas: um nível graduado de estado de conservação com um `condition_score` de 1 a 5; indicadores e tipos de vista; indicadores de reforma de cozinha e banheiro; piso de madeira e outros revestimentos; uma lista e uma contagem de melhorias; vagas e área de garagem; piscina, spa e outras características externas; contagem de lareiras; edificações acessórias (galpão, celeiro, oficina, casita, casa de hóspedes, canil etc.) com tamanhos extraídos de dimensões como “40x50”; estacionamento e depósito para motorhome; propriedade versus arrendamento de sistema solar e armazenamento em baterias; tamanho do lote (incluindo conversão de acre para pé quadrado); valores ligados ao contexto da venda associados a um avaliando (redução de preço, taxas de condomínio, arrendamento ou quitação do sistema solar, Mello-Roos); indícios de localização; e um escore de sentimento de destaques/preocupações em $[-1, 1]$.

15.2 Requisitos: vProlog (Trealla), e não SWI-Prolog

A interpretação das observações roda sobre o pacote R **vProlog** — um motor Trealla Prolog autocontido. Ele não está no CRAN; instale-o a partir do repositório r-universe do autor, como mostrado em “Instalação do earthUI e do vProlog”, no início deste artigo (binários pré-compilados para Windows e macOS, código-fonte em github.com/wcraytor/vProlog). Nenhuma instalação de Prolog no sistema (SWI-Prolog ou outra) é exigida ou utilizada. Se o vProlog estiver ausente, a interpretação das observações é pulada com um aviso, enquanto a divisão de `list` ainda é executada. O auxiliar `vProlog_available()` informa se o motor e a gramática incluída estão ambos presentes.

15.3 Habilitando e executando

1. Abra **Settings** (engrenagem) e marque **Enable Prolog / list column processing**. O seletor vem desligado por padrão, traz uma observação “advanced feature — for users with Prolog expertise”, é lembrado por projeto e é exibido na caixa de diálogo Project Information.
2. Na tabela de preditores, defina o papel **Special** das colunas relevantes como `remarks` ou `list` (várias colunas podem ter cada papel).
3. Use a seção **Execute Prolog Processing**, situada entre as Seções 4 e 5 da barra lateral. Ela oferece duas etapas:

1. **Expand lists & remarks** — executa a gramática DCG sobre cada coluna `remarks` e divide em one-hot cada coluna `list`. A expansão é *incremental*: apenas as colunas designadas desde a última execução são processadas, e o botão fica desabilitado quando tudo o que foi designado já tiver sido expandido. Uma barra de progresso acompanha a análise.

2. **Execute derivation rules** — aplica suas próprias regras Prolog `derive/2` sobre as colunas expandidas (veja abaixo). As regras passam por verificação de sintaxe e por validação em execução simulada antes de rodar.

15.4 Regras de derivação

Para cada linha de dados, o earthUI declara as colunas da linha como fatos `cell(Name, Value)` e coleta todas as soluções de suas regras `derive(Col, Val)` — cada `Col` distinto torna-se uma nova coluna, e o *primeiro* valor encontrado por coluna prevalece, de modo que uma cláusula alternativa pode vir em seguida. Predicados auxiliares (`count_true/2`, `count_false/2`, `count_true_prefix/2`, `cell_or/3`, `is_na/1`) dão suporte a agregação e ao tratamento de NA, e fatos `drop/1` descartam colunas depois que as regras rodam. Por exemplo:

```
% Merge a collinear cluster of solar flags into one boolean.
derive(solar_any, B) :-
  ( cell(pr_has_solar, true)
  ; cell(pr_solar_owned, true)
  ; cell(pr_has_battery, true) ) -> B = true ; B = false.

% Count amenities instead of OR-ing them (numeric column).
derive(amenity_count, N) :- count_true_prefix(pr_has_, N).

% Threshold -> boolean, with a fallback clause.
derive(big_outbuilding, true) :- cell(pr_outbuilding_sqft, S), S >= 1000.
derive(big_outbuilding, false).

% Discard a raw granular column once consolidated.
drop(pr_price_solar_lease).
```

As regras ficam em um arquivo `<project>_rules.pl` por projeto, criado a partir de um modelo integralmente comentado, e são editadas por **Edit derivation rules...** em um editor de código flutuante e arrastável, que também pode ser acoplado à barra lateral. Um botão **Check** valida as regras (varredura de sintaxe, verificação de carga e uma execução simulada contra a linha 1 dos seus dados) sem executá-las.

15.5 Arquivo de dados ativo versionado

Depois de qualquer uma das etapas, o data frame aumentado é gravado na pasta de entrada do projeto como `<original_name>_<YYMMDD_HHMMSS>.xlsx` e torna-se o arquivo de dados **ativo** do projeto — reabrir o projeto carrega os dados aumentados, de modo que as expansões nunca são recalculadas, e esse mesmo arquivo é o que o `glmnetUI` e o `mgcvUI` consomem. As colunas de origem das observações brutas são automaticamente mantidas fora do modelo; as colunas geradas *não* são incluídas automaticamente como preditores. A análise de texto pode produzir muitas colunas, portanto marque como **Latent** as que você (ainda) não quiser na regressão, ou desabilite-as com `×` (Capítulo 6).

15.6 Interface programática e arquivos incluídos

Todo o pipeline é exportado para uso em lote: `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` orquestra as passagens, construído sobre `remarks_to_columns()`, `split_list_column()`, `apply_user_rules()` e `validate_rules()`. Os fontes Prolog incluídos em `inst/prolog/` são `mls_remarks.pl` (a gramática DCG e a agregação de características), `mls_analyze.pl` (um driver de E/S de arquivos para análise em lote), `mls_rules.pl` (os predicados auxiliares disponíveis às regras de derivação) e `climate_regions_ca.pl` (abaixo).

15.7 Experimento relacionado: prioris por região climática

Um recurso experimental complementar trata das características com *dados escassos* — aquelas que a regressão não consegue estimar por aparecerem em apenas um punhado de comparáveis (um climatizador evaporativo, digamos). Três auxiliares em R — `climate_region_for()`, `market_area_profile()` e `climate_feature_priors()` — classificam a localização de um projeto (município + cidade) em uma de sete regiões climáticas da Califórnia reconhecidas pelos compradores (uma redução prática das 16 zonas climáticas de edificação da California Energy Commission) e retornam prioris ordinais de valor contributivo para características de climatização e resfriamento (aquecimento, ar-condicionado, climatizador evaporativo, exaustor de casa inteira, filtragem) por região. O módulo Prolog `climate_regions_ca.pl` espelha a mesma classificação para uso dentro das regras de derivação; ambos os lados são gerados de uma única fonte canônica em R, de modo que não podem divergir. Os prioris destinam-se a *informar* a valoração de características e a servir de verificação de plausibilidade dos coeficientes estimados — as atribuições de região (em particular no caso de municípios que abrangem várias regiões) devem ser revisadas pelo avaliador.

15.7.1 Além da Califórnia: pacotes de jurisdição e módulos de região de mercado

A Califórnia é a primeira jurisdição, não o limite. O `earthUI` é usado internacionalmente, e o conhecimento de prioris climáticos está sendo reestruturado como **pacotes de jurisdição**: tabelas de fatos autocontidas e com formato de dados (definições de zona, atribuições de unidades administrativas, prioris de características) atreladas aos códigos de país e de nível administrativo do sistema de projetos, de modo que um pacote possa descrever um estado dos EUA, um Land alemão ou um país pequeno inteiro. Uma camada complementar de **módulos de região de mercado** capta o vocabulário local e prioris refinados para uma região de mercado nomeada.

Ambos foram concebidos para serem elaborados *em campo* por Engenheiros de Avaliação e avaliadores, em seu próprio idioma — e, realisticamente, com assistência de LLM: o mundo não tem muitos programadores Prolog, de modo que os artefatos de campo são deliberadamente tabelas, e não programas. O formato do pacote é um esquema documentado com um exemplo completo resolvido (a própria Califórnia), expressamente para que um assistente como o Claude possa redigir um pacote válido a partir de uma descrição de mercado em linguagem comum; o earthUI gera o Prolog subjacente a partir do pacote, e um validador bloqueante (verificações de esquema, cobertura, integridade referencial e execução simulada) precisa passar antes que um pacote possa ser habilitado. Todo pacote elaborado em campo carrega procedência obrigatória — o autor responsável, a data e a autoridade citada — de modo que o profissional responsável, e não a ferramenta de redação, respalda o conhecimento, e os relatórios podem citar exatamente o pacote e a versão utilizados.

Dica: Um padrão prático: expanda as observações, consolide os indicadores granulares `pr_*` em um punhado de contagens ou booleanos derivados com regras `derive/2`, aplique `drop/1` aos indicadores brutos que você consolidou e mantenha apenas as colunas derivadas como preditores candidatos. Isso mantém a lista de preditores curta e a colinearidade administrável.

Referências

- Friedman, Jerome H. (1991). “Multivariate Adaptive Regression Splines”. Em: *The Annals of Statistics* 19.1, pp. 1–141. DOI: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- Milborrow, Stephen (1 de out. de 2024). *earth: Multivariate Adaptive Regression Splines*. R package. URL: <https://CRAN.R-project.org/package=earth>.
- (s.d.[a]). *Notes on the earth Package*. URL: <http://www.milbo.org/doc/earth-notes.pdf>.
- (s.d.[b]). *Variance Models in earth*. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.