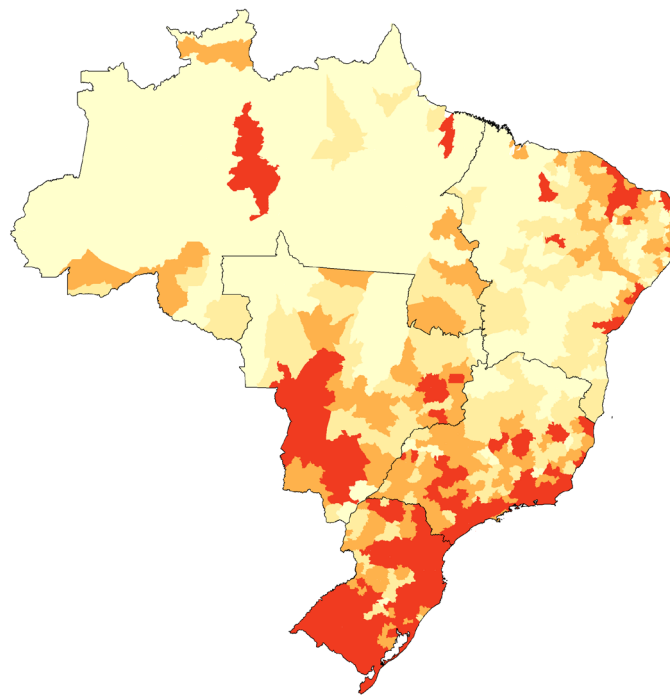


Figura 1. Taxa de mortalidade por câncer de mama da população feminina nas regiões geográficas imediatas brasileiras por quartis entre 2015 e 2019. As regiões de cor creme, mais claras, apresentaram as menores taxas, enquanto as de cor vermelha mostraram as maiores. Os valores dos quartis foram calculados dentro da própria variável.

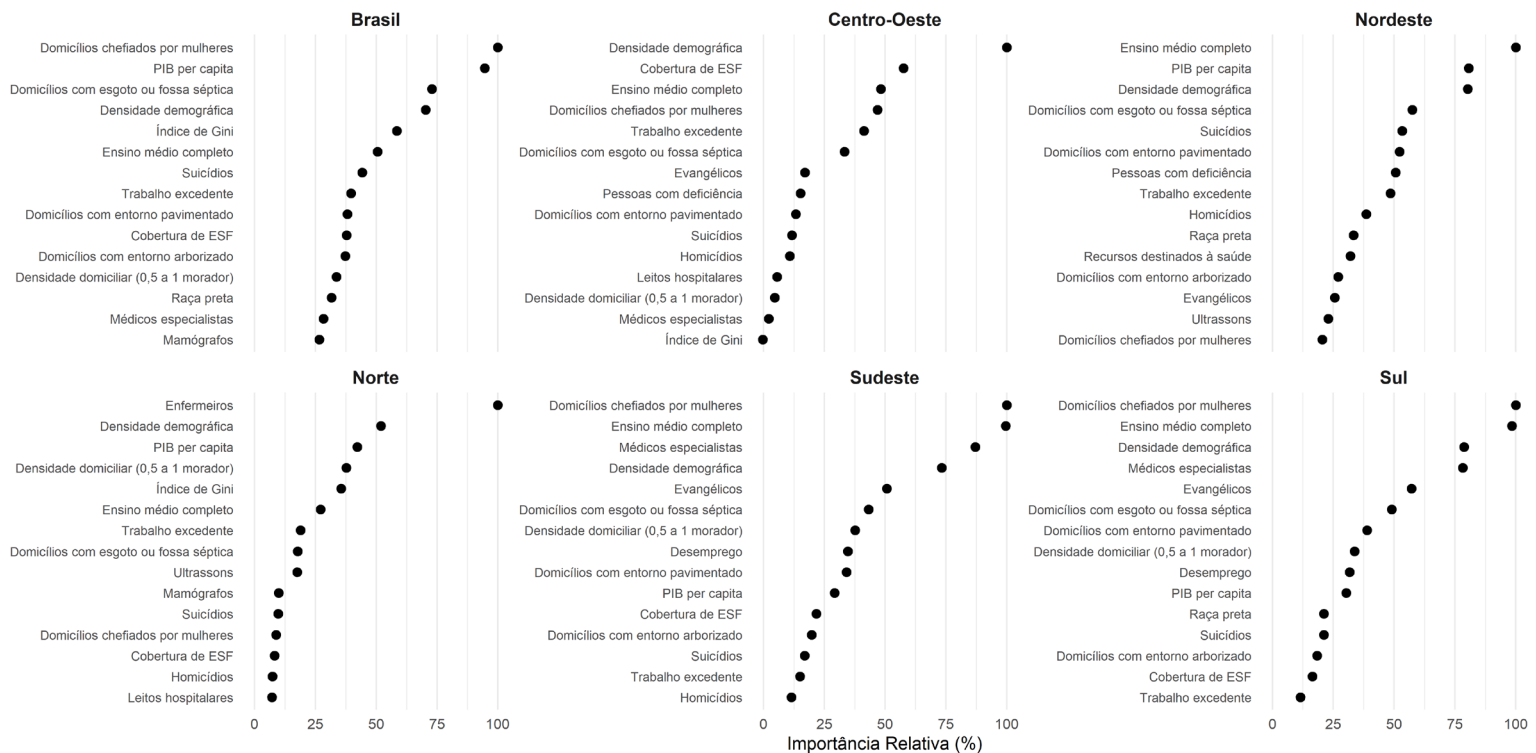


Figura 2. Importância relativa das 15 variáveis socioeconômicas e de saúde que apresentaram os maiores valores de importância na modelagem Random Forest para o Brasil e cada uma das grandes regiões – Norte, Nordeste, Centro-Oeste, Sudeste e Sul. A medida considerada neste gráfico foi o IncMSE% (aumento no erro quadrático médio) ao permutar-se cada variável. Isso é, se quando uma variável tem o seu valor original alterado, o desempenho do algoritmo é consideravelmente afetado, ela tem importância relativa relevante para um bom desempenho do modelo. O valor do IncMSE% foi escalonado entre 0% e 100% (maior importância relativa) para facilitar a interpretação dos resultados.

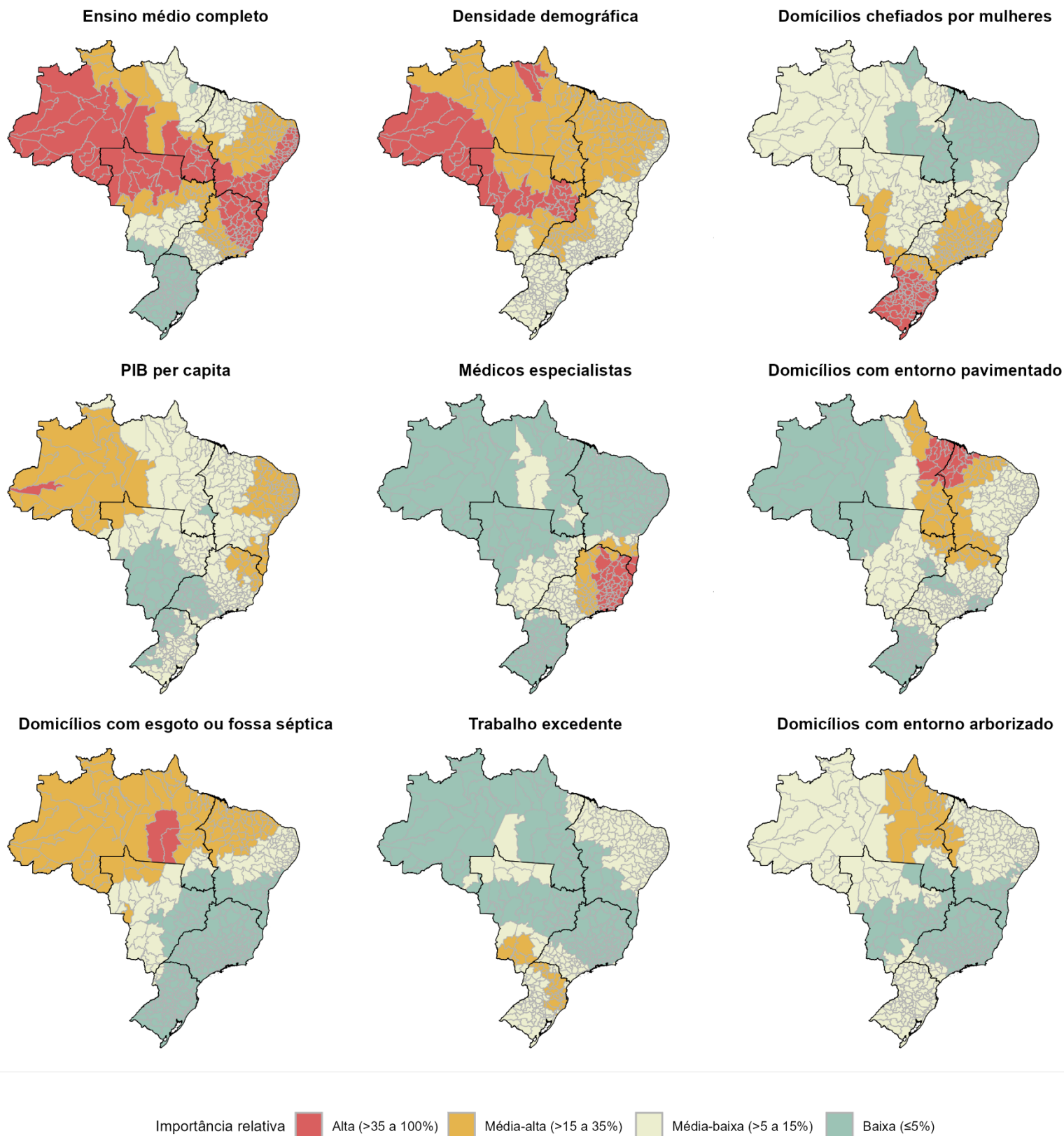


Figura 3. Mapas de importância relativa local das 9 variáveis que tiveram maiores médias de importância no desempenho no modelo de Random Forest Geográfico. As linhas pretas delimitam as grandes regiões brasileiras (Norte, Nordeste, Centro-Oeste, Sudeste, Sul). A medida considerada neste gráfico foi o IncMSE% (aumento no erro quadrático médio) ao permutar-se cada variável. O valor mais alto dentre todas as importâncias das variáveis foi referenciado como 100%. Todos os outros valores foram escalonados relativamente ao valor mais alto. As últimas duas variáveis (“Trabalho excedente” e “Domicílios com entorno arborizado”) não apresentaram importância relativa alta em nenhuma das 510 regiões geográficas imediatas brasileiras.

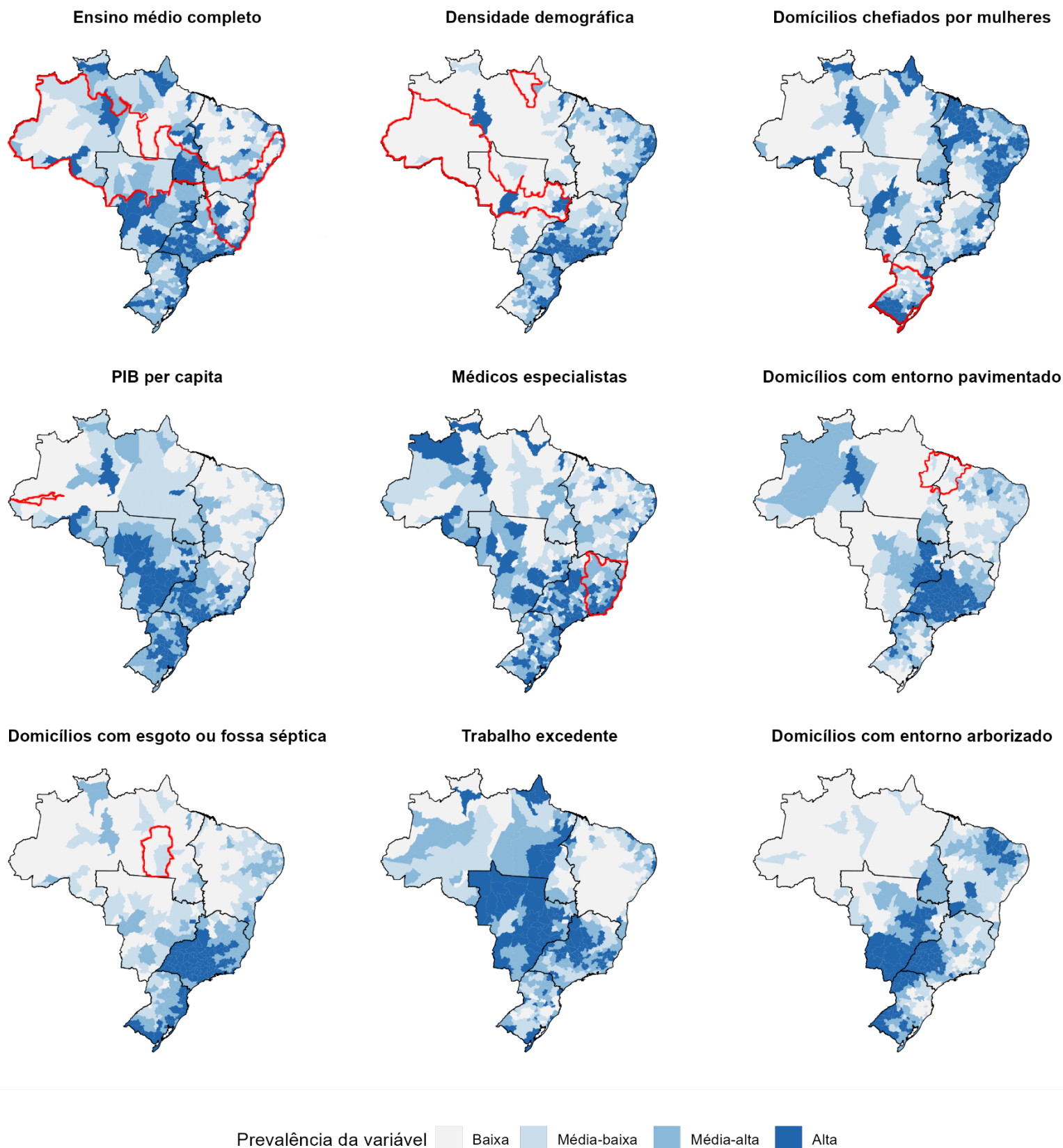


Figura 4. Mapa de prevalência dos fatores de risco e delimitação das áreas que apresentaram alta importância da variável na análise de Random Forest Geográfico. Todas as variáveis usaram as suas respectivas escalas e foram divididas nos quartis. A delimitação das áreas de alta importância foi realizada na cor vermelha. As linhas pretas delimitam as grandes regiões brasileiras (Norte, Nordeste, Centro-Oeste, Sudeste, Sul). As últimas duas variáveis (“Trabalho excedente” e “Domicílios com entorno arborizado”) não apresentaram importância relativa alta em nenhuma das 510 regiões geográficas imediatas brasileiras.