

Projeção a quantidade de passageiros no transporte público de Fortaleza

Resumo:

As decisões de uma empresa dependem fundamentalmente das estimativas dos dados por ela gerados. Quanto mais precisas forem essas estimativas, menor a possibilidade de erros serem cometidos.

No caso do transporte público de Fortaleza, erros podem gerar superlotações nos ônibus, rotas mal planejadas e desequilíbrio orçamentário. Nesse estudo, apresentamos dois modelos preditivos para a previsão dos primeiros 2 meses do ano de 2020, com base no estudo do quantitativo de passageiros dos últimos 20 anos.

Modelos utilizados:

A suavização exponencial e os modelos ARIMA são as duas abordagens mais amplamente usadas para previsão de séries temporais e fornecem abordagens complementares para o problema. Embora os modelos de suavização exponencial sejam baseados em uma descrição da tendência e da sazonalidade nos dados, os modelos ARIMA visam descrever as autocorrelações nos dados.

Total de Passageiros em Fortaleza nos últimos 20 anos

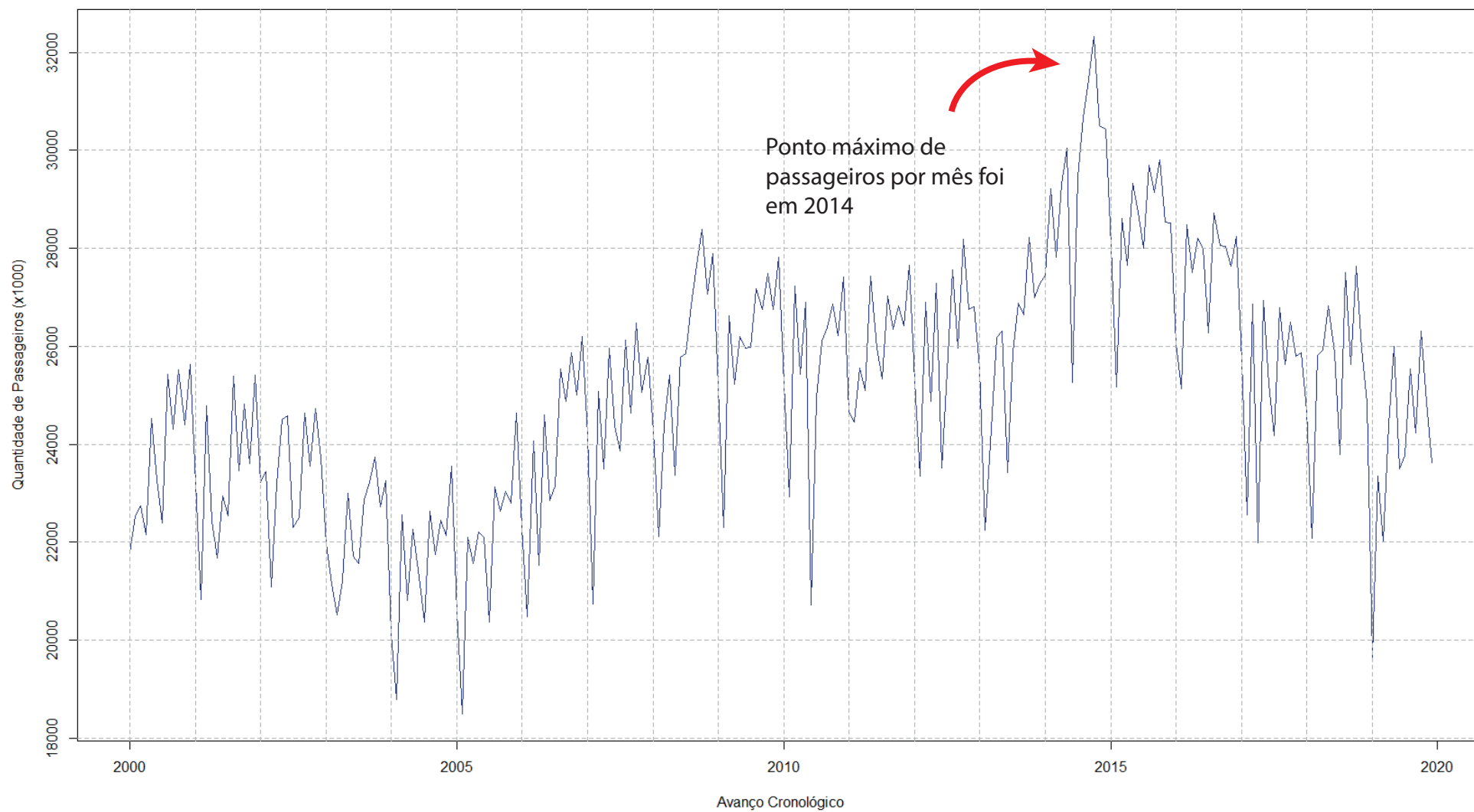
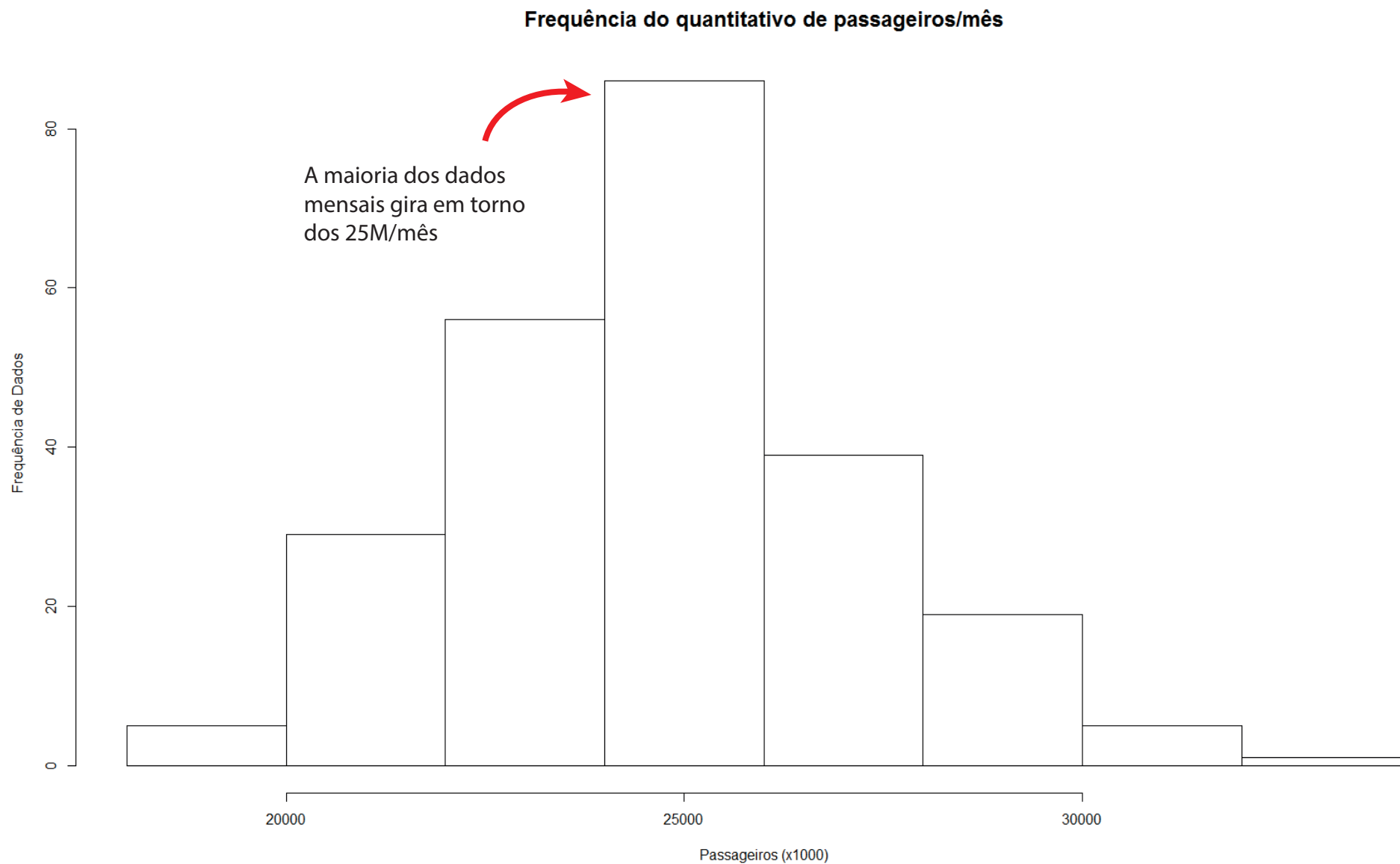
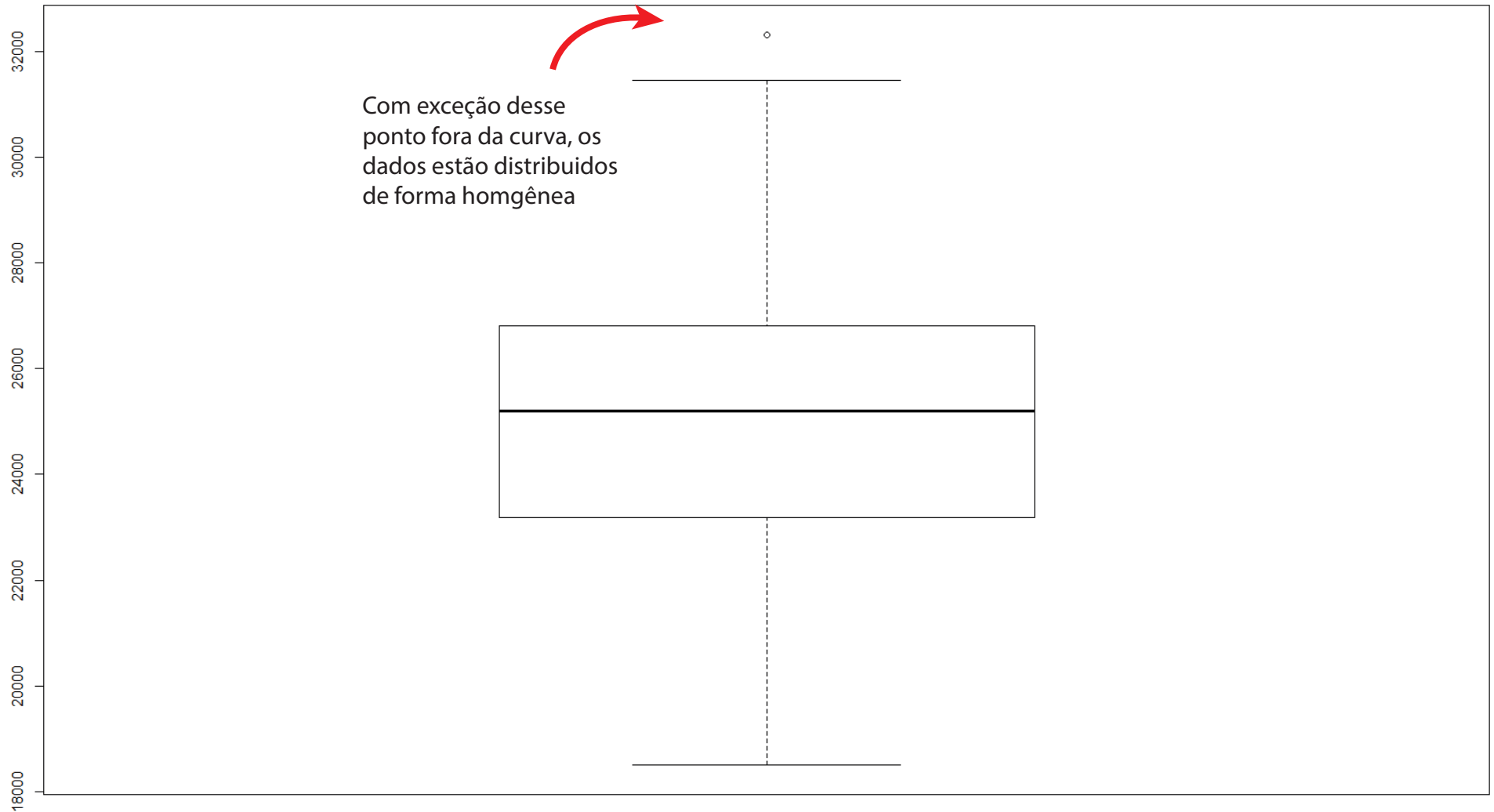


Gráfico aritmético simples para demonstração de quantitativo absoluto



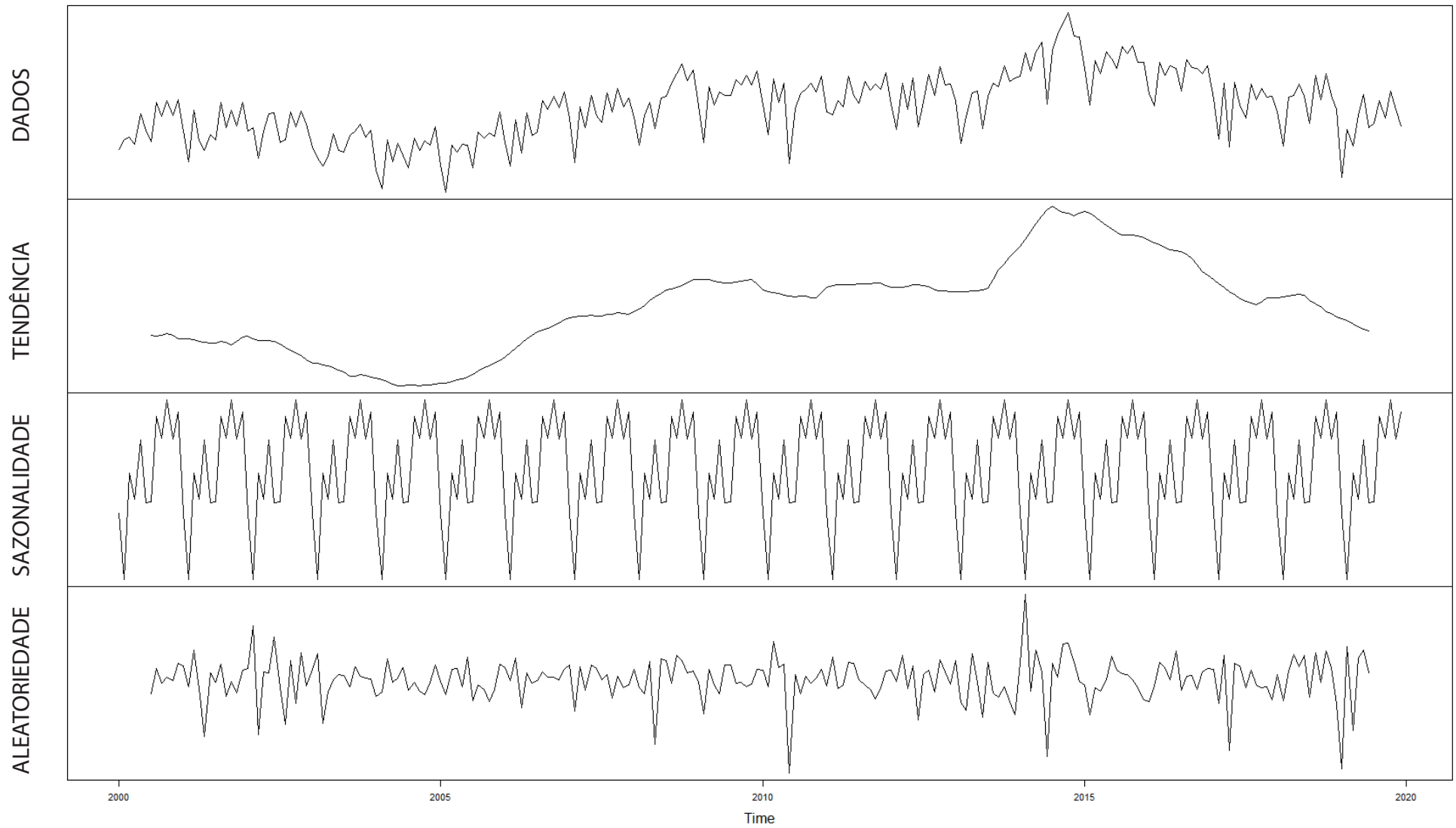
O histograma dispõe as informações de modo que seja possível a visualização da forma da distribuição de um conjunto de dados e também a percepção da localização do valor central e da dispersão dos dados em torno deste valor central.

Distribuição dos dados de passageiros nos últimos 20 anos

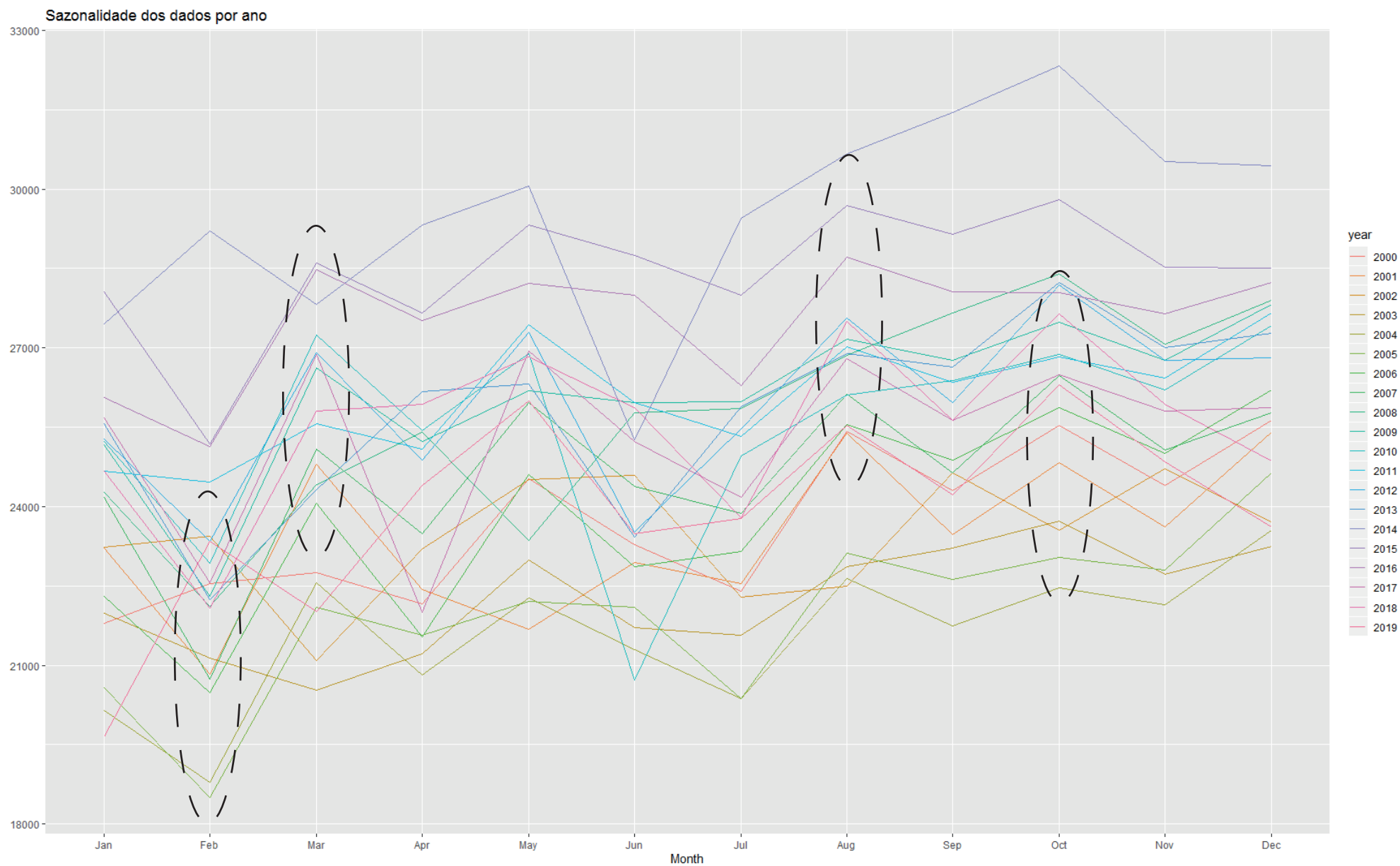


O boxplot tem como objetivo estudar as medidas estatística do conjunto de dados, como propriedades de locação, variabilidade, média, e outliers (pontos fora da curva). No boxplot a parte central do gráfico contém os valores que estão entre o primeiro quartil e o terceiro quartil. As hastes inferiores e superiores se estendem, respectivamente, do primeiro quartil até o menor valor, limite inferior, e do terceiro quartil até o maior valor.

Decomposição de Dados



Decomposição de dados, com a finalidade de detectar com base na totalidade dos dados, a tendência de crescimento ou queda, a sazonalidade dos dados, e se há aleatoriedade no conjunto coletado



Através desse gráfico é possível detectar a sazonalidade, localizando pontos onde há um padrão de alta ou baixa nos dados

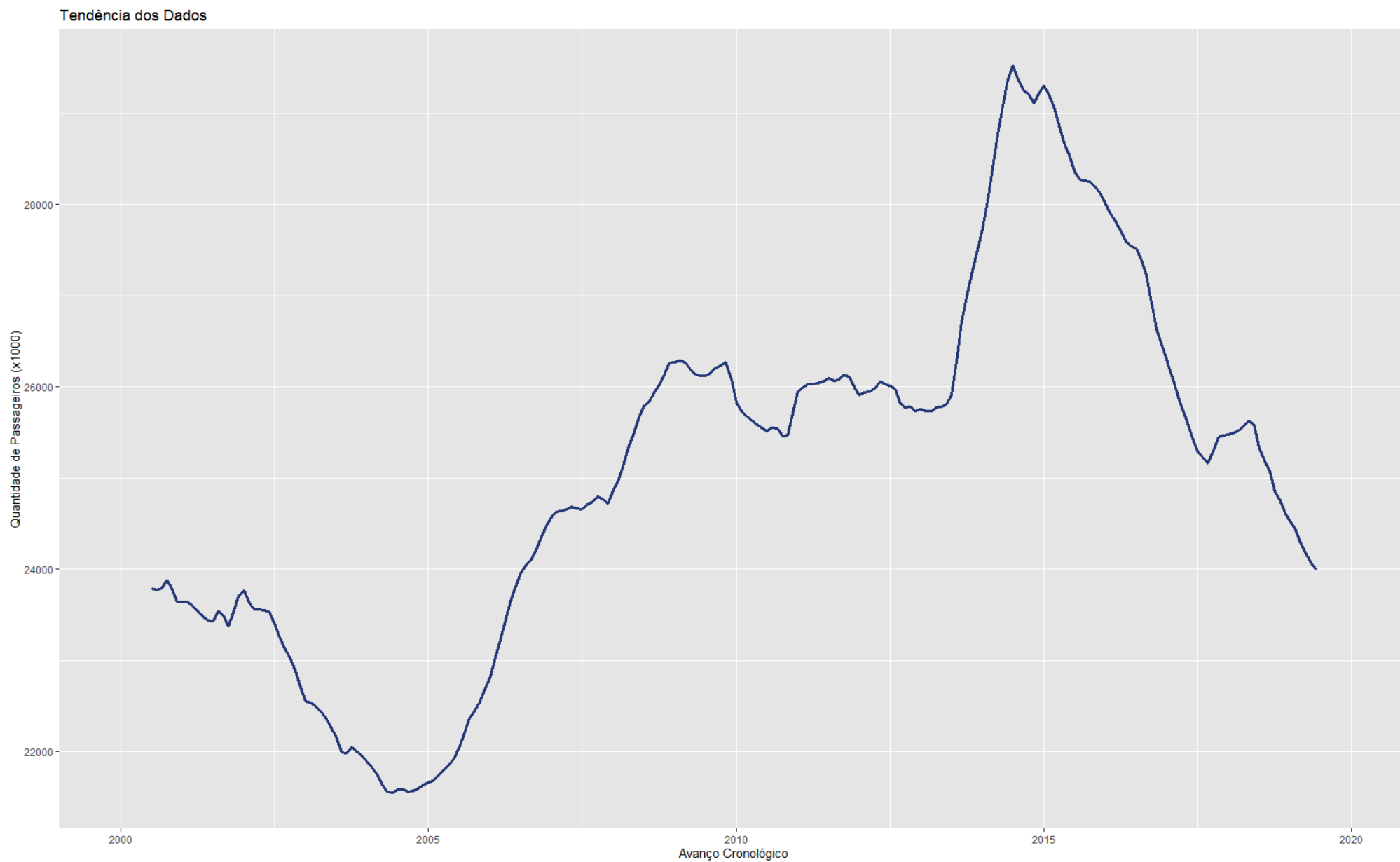
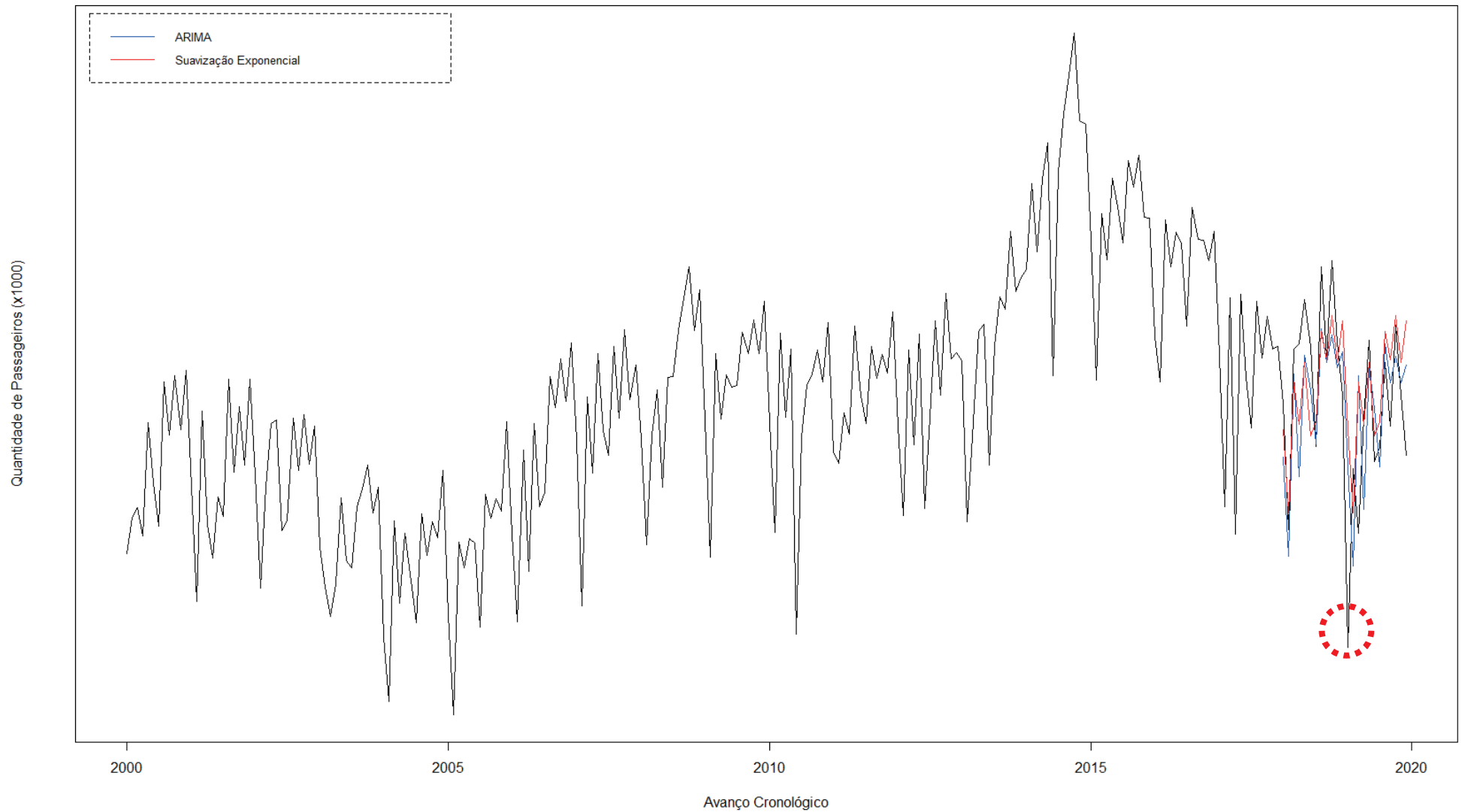
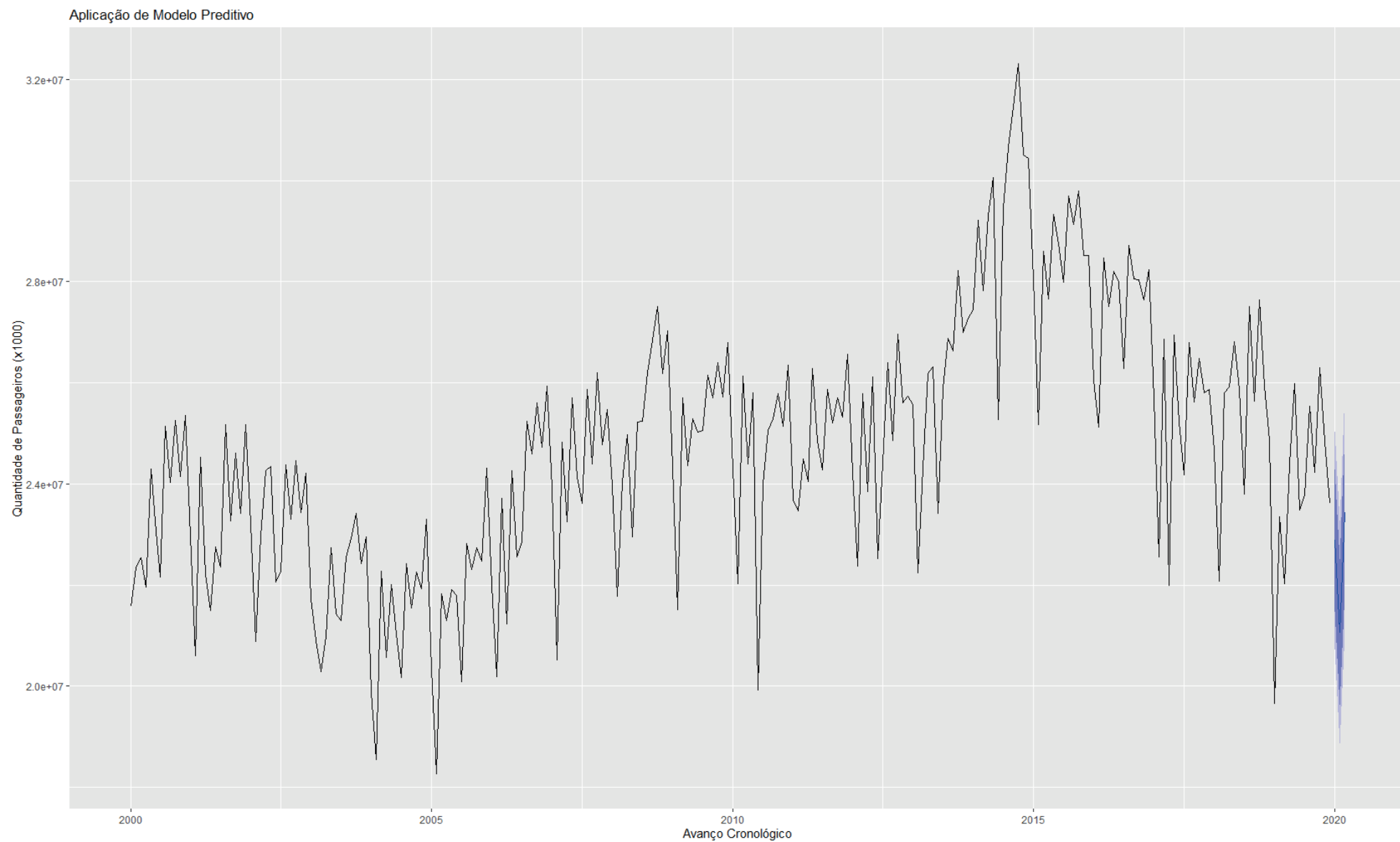


Gráfico aritmético absoluto padrão, mostrando a tendência dos dados isolada do conjunto total

Teste dos Modelos Preditivos



Com base nos dados fornecidos, testamos dois modelos preditivos para o conjunto de dados, que levassem em consideração as nuances fornecidas. Os resultados foram bem semelhantes, com apenas algumas discrepâncias, como por exemplo, a queda em janeiro de 2019 (em destaque) por conta dos ataques aos ônibus da cidade, o que causou uma queda na circulação de transporte.



Verificando a eficácia do modelo, ele foi aplicado para os três primeiros meses de 2020.

| | Dados Reais | Dados Previstos | Variação |
|------------------|------------------------|----------------------------|-----------------|
| Janeiro | 23002126 | 22537732 | 2.02% |
| Fevereiro | 21174990 | 21071959 | 0.49% |

Conclusão

Concluimos que os modelos testados mostram eficácia, comparados aos dados coletados nos dois primeiros meses de 2020. Mesmo com a pandemia do Coronavírus, quando corretamente aplicada, a tendência de dados pode dar uma estimativa de passageiros durante esse período no transporte público de Fortaleza