

Visualização de dados: Uma abordagem baseada no {ggplot2}

Conhecendo o mercado aéreo brasileiro

Prof. Wagner Hugo Bonat

Projeto prático: Conhecendo o mercado aéreo brasileiro

Considere o conjunto de dados sobre voos e operações aéreas disponibilizado pela ANAC (Agência Nacional de Aviação Comercial) no seguinte endereço clique aqui. O arquivo chamado de `Dados_Estatisticos.csv` apresenta uma série de informações sobre as operadoras aéreas que atuam no Brasil. Forneça o código R usando o `framework` {tidyverse} para responder cada uma das seguintes questões práticas. Para a entrega da lista crie um arquivo .Rmd com as perguntas, o código R e os gráficos associados. Envie o arquivo .html gerado para correção.

1. Considere que é de interesse visualizar a série temporal (anual) do número de decolagens para as 10 empresas nacionais que mais decolaram no período disponível de dados. Faça duas visualizações uma para enfatizar a comparação entre as empresas e outra para enfatizar a visualização da série de cada empresa. Considere apenas os registros que tenham o número de decolagens.
2. Faça uma visualização de dados comparar o número de passageiros pagantes entre empresas nacionais e estrangeiras ao longo dos anos.
3. Faça uma visualização de dados comparar o número de passageiros pagantes entre empresas nacionais e estrangeiras de acordo com o grupo de voo.
4. Faça uma visualização para enfatizar a distribuição da variável número de passageiros pagantes de acordo com a região de origem e natureza do voo. Enfatize a comparação entre DOMÉSTICA e INTERNACIONAL para cada região de origem. Para criar o gráfico use o log10 da variável PASSAGEIROS_PAGOS.
5. Crie uma variável que indique a eficiência no uso do combustível como sendo a razão entre o DISTANCIA_VOADA_KM e COMBUSTIVEL_LITROS. Compare as empresas de acordo com a região de origem e destino. Use a escala y em log10 para facilitar a comparação das regiões.
6. Faça uma visualização para mostrar a relação entre o consumo de combustível e a distância voada. Considere empresas nacionais e internacionais e os diferentes anos para enriquecer a visualização. Considere transformar as variáveis e use alguma forma estatística para enfatizar a forma do relacionamento.
7. Baseado na visualização em 6 você diria que tem alguma inconsistência com esses dados? Justique.
8. Considerando apenas as empresas nacionais. Faça uma visualização para investigar se a região do aeroporto de origem influencia na relação entre combustível e distância voada.
9. O indicador RTK (Revenue tonne kilometer) refere-se ao volume de toneladas quilômetros transportada. Faça uma visualização da relação deste indicador com o consumo de combustível. Filtre apenas observações com ambas variáveis maiores que zero. Considere também que é de interesse comparar essa relação entre os diferentes anos. Considere incluir alguma camada estatística para enfatizar o formato do relacionamento entre as variáveis.
10. O indicador ATK (Available tonne kilometer) refere-se ao volume de tonelada quilômetro oferecida pela companhia. Crie uma visualização para explorar a relação entre estes indicadores. Considere aspectos como GRUPO_DE_VOOS e NATUREZA além dos anos.