

PCC170 - Projeto e Análise de Experimentos Computacionais

Marco Antonio M. Carvalho

Departamento de Computação
Instituto de Ciências Exatas e Biológicas
Universidade Federal de Ouro Preto



1 Representações gráficas – parte 1

2 Boxplot

- Gráficos violino

Fonte

Este material é parcialmente baseado no conteúdo de

- ▶ FARIAS, A. M. L.; ALVES, R.M. 2010. Conteúdos digitais para o ensino e aprendizagem de matemática e estatística: *Construindo boxplots*. Instituto de Matemática, Universidade Federal Fluminense.
- ▶ PARREIRA, Guilherme. *Gráfico boxplot*. 2018. Disponível em <https://bityli.com/WTJWuY>.

Licença

Este material está licenciado sob a Creative Commons BY-NC-SA 4.0. Isto significa que o material pode ser compartilhado e adaptado, desde que seja atribuído o devido crédito, que o material não seja utilizado de forma comercial e que o material resultante seja distribuído de acordo com a mesma licença.

Introdução

Representações gráficas são particularmente úteis para apresentar a comparação, distribuição, relação ou composição de um conjunto de dados.

Diferentes softwares, como *R* (biblioteca *ggplot2*^a) e *gnuplot*, e bibliotecas como *matplotlib* podem ser utilizados para criar gráficos vetoriais com estilo científico.

Gráficos devem ser utilizados de maneira complementar em relação a tabelas e ao próprio texto e **jamais** de maneira redundante.

Um guia rápido para escolha de gráficos pode ser encontrado em <https://bityli.com/momen>.

^aVeja uma galeria de exemplos em <https://r-graph-gallery.com>

Representações gráficas

Gráficos de comparação

Utilizados para confrontar dados ao longo do tempo ou entre várias categorias.

Por exemplo, confrontar os indicadores de diferentes algoritmos em diferentes conjuntos de instâncias.

Inclui gráficos de colunas, barras, radar e linhas.

Comparação:



fonte: <https://bityli.com/FWqRW>

Representações gráficas

Gráficos de distribuição

Utilizados para mostrar a frequência em que ocorrem os dados.

Por exemplo, *gap* ou desvio padrão no valor das soluções para um conjunto de instâncias.

Inclui histogramas, gráficos de superfície e de dispersão.

Distribuição:



fonte: <https://bityli.com/FWqRW>

Representações gráficas

Gráficos de relação

Utilizados para mostrar a interdependência entre variáveis.

Por exemplo, ilustrar objetivos conflitantes ou não.

Inclui gráficos de dispersão e de bolhas.

Relação:



fonte: <https://bityli.com/FWqRW>

Representações gráficas

Gráficos de composição

Utilizados para mostrar a participação dos componentes de um todo.

Por exemplo, os componentes do valor da solução (FO, penalidades, etc.) ou os componentes do tempo de execução (solução inicial, cada busca local, etc.).

Inclui gráficos de colunas empilhadas, área empilhada, pizza^a e cachoeira.

^aapenas para 2 variáveis. Para mais variáveis, utilize gráficos de barras.

Composição:



fonte: <https://bityli.com/FWqRW>

Outros gráficos

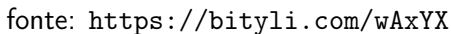
Para além dos gráficos apresentados anteriormente, veremos detalhes sobre gráficos especificamente úteis para utilizarmos em nossos experimentos:

- ▶ *Boxplots*, ou diagramas de caixa;
- ▶ Gráficos de coordenadas paralelas;
- ▶ Gráficos *time-to-target*;
- ▶ Gráficos múltiplos *time-to-target*.

Boxplot

Um *boxplot*, ou *diagrama de caixa*, é um gráfico utilizado para avaliar a distribuição empírica de dados, e é poderoso por concentrar muita informação com simplicidade.

Pode ser utilizado também para uma comparação visual entre dois ou mais grupos de dados, como os resultados obtidos por dois algoritmos.



Boxplot

De maneira sucinta, um *boxplot* resume um conjunto de dados em cinco valores:

- ▶ Valor mínimo;
- ▶ Primeiro quartil (Q1);
- ▶ Segundo quartil (Q2);
- ▶ Terceiro quartil (Q3).
- ▶ Valor máximo.

Adicionalmente, valores *outliers*, ou discrepantes, são também exibidos.

Quartis

Os três quartis de um *boxplot* dividem os dados da seguinte maneira:

Primeiro quartil: 25% dos dados estão abaixo dele;

Segundo quartil: 50% dos dados estão abaixo dele, representa a mediana;

Terceiro quartil: 75% dos dados estão abaixo dele.

A caixa do *boxplot* é construída desde o primeiro quartil até o terceiro, paralela ao eixo da escala dos dados.

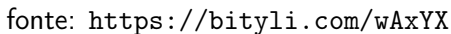
Valores mínimo e máximo

Os valores **mínimo** e **máximo** de um conjunto de dados são conectados à caixa do *boxplot* por uma linha, que vai de cada extremidade da caixa ao correspondente valor extremo dos dados.

Valores discrepantes (*outliers*)

Valores discrepantes, ou *atípicos*, são valores muito afastados da grande parte dos dados, i.e., muito abaixo de Q1 ou muito acima de Q3.

Estes valores são representados individualmente, acima ou abaixo do *boxplot*.



Boxplots

Os *boxplots* podem ser utilizados para representar de maneira poderosa a maior parte dos indicadores de desempenho vistos anteriormente:

- ▶ Valor da função objetivo;
- ▶ Tempo de execução;
- ▶ Indicadores relacionados aos atributos específicos das instâncias para um problema.

Entretanto, é necessário estar atento para não misturar dados não relacionados entre si em um mesmo *boxplot*.

Boxplots

Os exemplos a seguir foram extraídos do artigo

- ▶ Gandra, V. M. S., Çalık, H., Wauters, T., Toffolo, T. A., Carvalho, M. A. M., & Berghe, G. V. (2021). *The impact of loading restrictions on the two-echelon location routing problem*. Computers & Industrial Engineering, 160, 107609.

Two-echelon location routing problem

Este é um problema de roteamento em dois níveis, em que itens vindos de diferentes origens são distribuídos a clientes utilizando instalações intermediárias.

Itens se encontram em um ou mais depósitos centrais e então veículos capacitados fazem a entrega para depósitos satélites (intermediários). De lá, outros veículos fazem a entrega para os clientes.

Idealmente, os itens devem estar carregados sequencialmente dentro dos veículos de acordo com a ordem das entregas. Porém, podem ser necessária a manipulação de itens entre as entregas, exceto rotações.

Representações gráficas

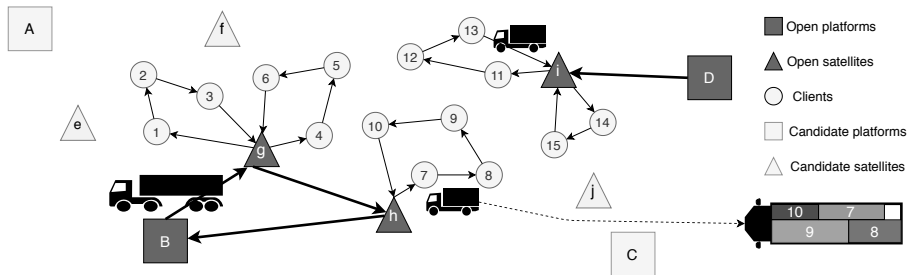


Figure 1: Two-echelon routes with load plan.

Two-echelon location routing problem

Os indicadores padrão são tempo computacional e valor da função objetivo.

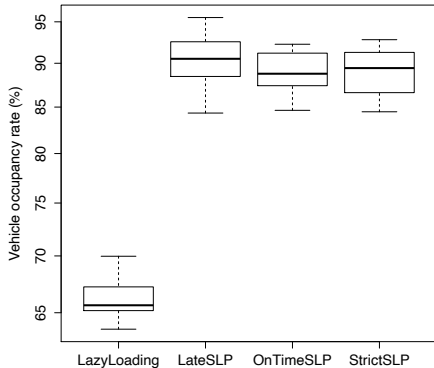
Os indicadores de atributos específicos incluem taxa de ocupação dos veículos, número de rotas geradas, duração das rotas e penalidade (em unidades de tempo) por reorganizar itens dentro dos veículos.

Two-echelon location routing problem

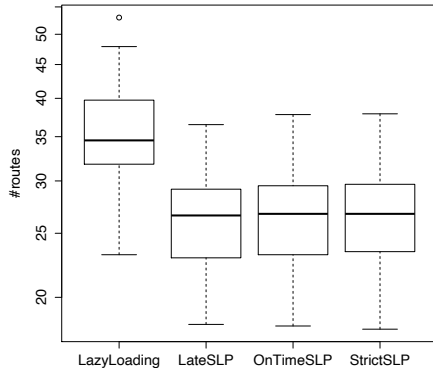
Quatro estratégias de carregamento de veículos foram propostas dentro de uma ILS e comparadas:

- ▶ LazyLoading: adiciona itens ao veículo enquanto houver área residual no veículo;
- ▶ LateSLP: gera rotas e só ao final adiciona as penalidades de carregamento;
- ▶ OnTimeSLP: gera rotas e informa as penalidades de carregamento a cada instante;
- ▶ StrictSLP: não gera rotas com penalidades por manipulação extra de itens.

Representações gráficas

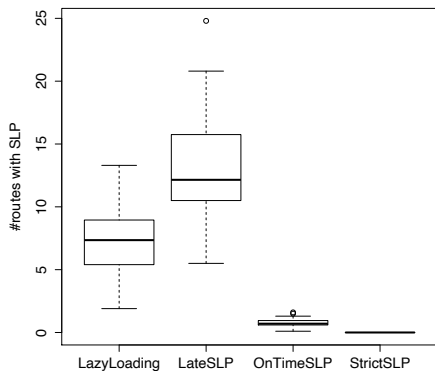


(a)

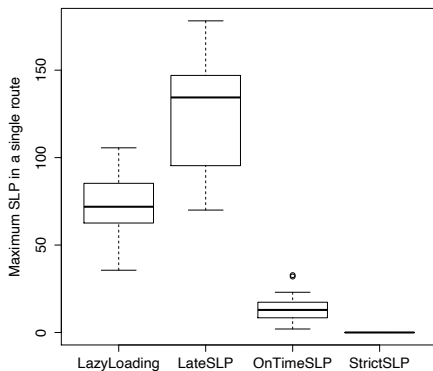


(b)

Representações gráficas



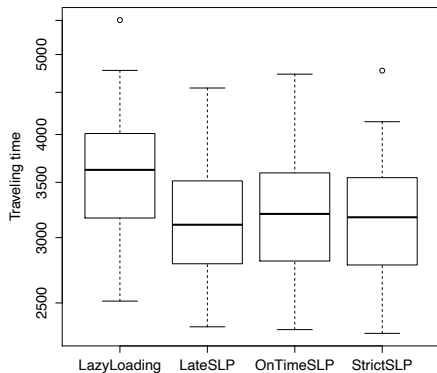
(c)



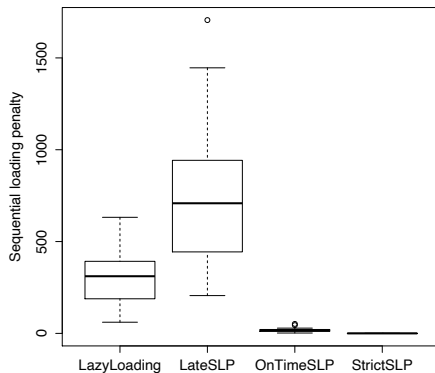
(d)

Figure 8: Different loading strategies for the second echelon.

Representações gráficas

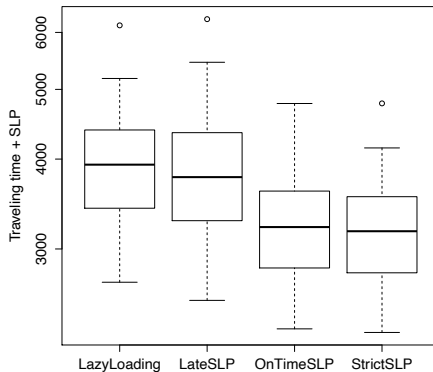


(a)



(b)

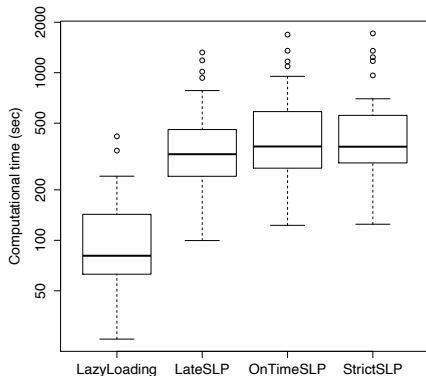
Representações gráficas



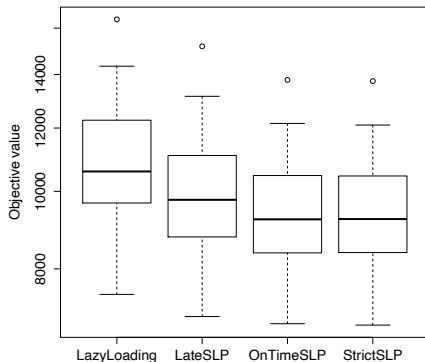
(c)

Figure 9: Different loading strategies for the second echelon. Traveling time and penalty analysis.

Representações gráficas



(a)



(b)

Figure 10: Best solution comparison and run time of different loading strategies.

Análise

Além de apresentar gráficos, é necessário incluir análises adicionais.

Note que isto é diferente de acrescentar texto “lendo” o que já está descrito nos gráficos, o que seria redundante.

É interessante analisar os fatos por trás dos dados descritos nos gráficos e apresentar dados adicionais.

Análise

- ▶ Porquê LazyLoading possui uma ocupação de veículos tão baixa (Fig. 8a) ou gera tantas rotas (Fig. 8b)?
- ▶ Qual a proporção entre o número de rotas geradas (Fig. 8b)?
- ▶ Qual o percentual de rotas com penalidades?
- ▶ Qual a relação entre ocupação de veículos e penalidades?
- ▶ Qual a proporção entre maior tempo de viagem e maior penalidade?
- ▶ Indicadores de desempenho;
- ▶ Testes estatísticos.

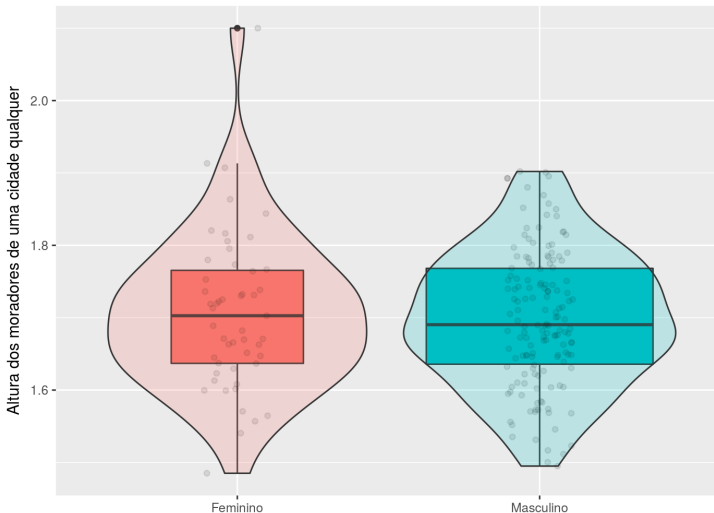
Definição

O “violino” são curvas no entorno de um gráfico boxplot.

Estas curvas representam a função densidade de probabilidade estimada: curvas mais largas representam maior densidade de pontos.

Esse tipo de representação pode ser útil para alguns conjuntos de dados, pois adiciona a dimensão de densidade.

Gráfico violino



fonte: <https://bityli.com/WTJWuY>

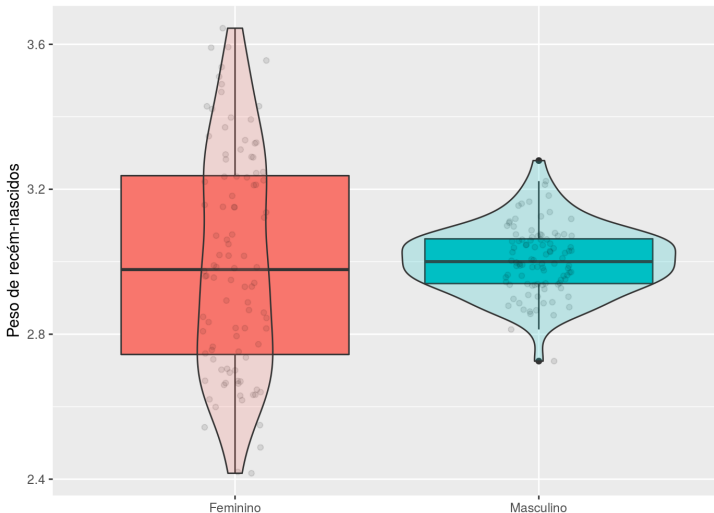
Exemplo 1

Note que para os indivíduos do sexo masculino, a forma do violino (densidade de pontos) é um pouco mais larga abaixo da mediana.

Já para os indivíduos do sexo feminino, a forma do violino está bem “arredondada”, ou densa no entorno da mediana.

Ou seja, os indivíduos do sexo feminino tiveram alturas levemente superiores as dos indivíduos do sexo masculino.

Gráfico violino



fonte: <https://bityli.com/WTJWuY>

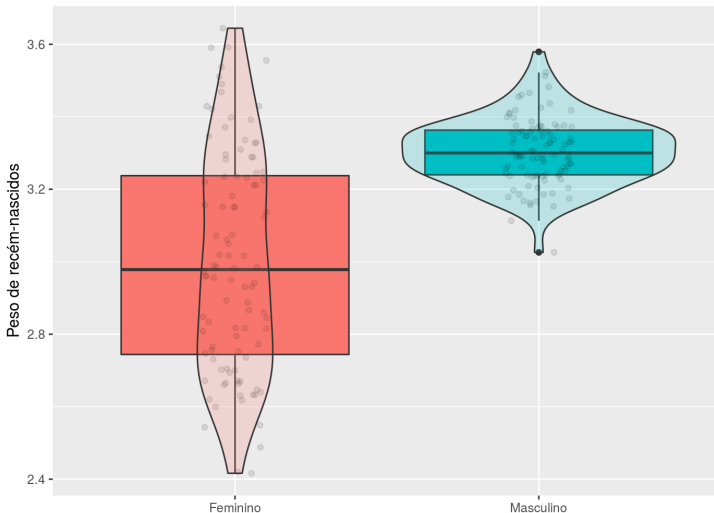
Exemplo 2

Nota-se que os recém-nascidos do sexo feminino variaram de cerca de 2,4kg até 3,6kg.

Os recém-nascidos do sexo masculino variaram por cerca de 2,8kg até 3,2kg.

Embora a variabilidade dos recém-nascidos do sexo feminino seja bem maior que os do sexo masculino, denotada pela alta dispersão, a mediana foi similar entre os grupos.

Gráfico violino



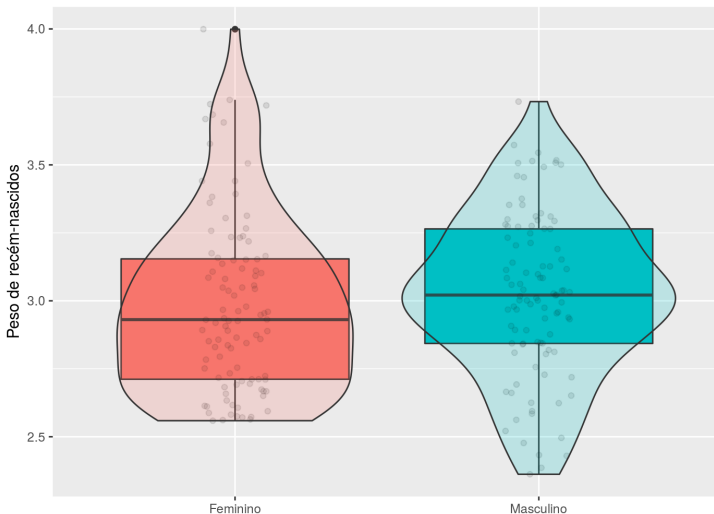
fonte: <https://bityli.com/WTJWuY>

Exemplo 3

Neste exemplo, foram mantidos os mesmos dados do gráfico anterior para o sexo feminino, e para o sexo masculino foi adicionado 0,3kg ao peso de cada recém-nascido.

Fica claro que os recém-nascidos do sexo masculino tiveram maior peso que os recém-nascidos do sexo feminino, além de uma menor variabilidade.

Gráfico violino



fonte: <https://bityli.com/WTJWuY>

Exemplo 4

Neste exemplo, manteve-se a mesma representação para o peso dos recém-nascidos para o sexo masculino. Para o sexo feminino, apenas para ilustração, supôs-se existir um maior número de recém-nascidos do sexo feminino com sobrepeso.

O peso mediano dos recém-nascidos para o sexo masculino foi igual a 3kg e a variabilidade ao redor da mediana se deu de forma proporcional, tanto acima e abaixo.

Para os recém-nascidos do sexo feminino, houve peso maior que 3,7kg, enquanto não houve peso menor que 2,5kg. Este é um indicativo de assimetria

A partir do violino, nota-se uma maior densidade de pontos entre 2,5kg e 3,25kg, e à medida em que afasta-se de 3,5kg, a densidade de pontos diminui.

Leitura recomendada

- ▶ DP6 Team. 2016. *Como escolher o melhor gráfico para meus dados?* — *DataViz Basics: 2 de 4*. Disponível em <https://bityli.com/FWqRW>.
- ▶ Kaushi, A. 2016. *Análise de dados: qual é o gráfico mais adequado?*. Disponível em <https://bityli.com/momen>.

Dúvidas?

