

Visualizações, problema 1

[Objetivos](#)

[O resultado](#)

[Antes de começar](#)

[Parte 1: Visualizações alheias](#)

[Parte 2: perguntas sobre 10 mil das músicas no mundo](#)

[Parte 3: visualizações em papel](#)

[Parte 4: Código](#)

Objetivos

- Praticar a crítica de visualizações como suporte a tarefas de análise usando a teoria que discutimos
- Criar perguntas de análise que possam ser examinadas via visualização
- Conceber visualizações úteis às tarefas escolhidas
- Implementar visualizações usando R e ggplot2

O resultado

O que esperamos como resultado deste problema é um relatório respondendo e comentando os pontos levantados. Lembre que o importante é comunicar suas conclusões e como você chegou nelas a partir dos dados. Procure ser preciso nas suas afirmações, e use os dados para embasar as afirmações.

Questões práticas:

- Seu resultado deverá estar em um repositório no github
- Use o [R Markdown](#) para fazer o relatório. Simplificará a sua vida.
- Você pode publicar o seu relatório usando o [Rpubs](#).

- Seu código provavelmente sairá do trivial para esse problema. Se isso acontecer:
 1. Coloque a definição de funções em um arquivo separado do markdown
 2. Faça testes de unidade para qualquer função que vá além do trivial. Queremos ter confiança que suas descobertas não são bugs no seu script.
-

Antes de começar

1. Crie um repositório no github para esta atividade e [coloque o link aqui](#)
-

Parte 1: Visualizações alheias

1. Antes de começar suas próprias visualizações, pratique a desconstrução de uma visualização de outra pessoa. Encontre e escolha uma visualização que use uma série temporal e/ou dados com mais de dois atributos (mas que não seja um mapa nem um grafo). Em seguida, poste essa visualização em uma nova thread no piazza (devemos ter uma thread por visualização).

No post, inclua uma descrição da pergunta sendo tratada, dos dados e de como cada atributo dos dados foi codificado. Justifique tecnicamente por que essa é uma boa visualização para a pergunta que você identificou. (Não diga que é uma boa visualização porque ela é eficiente/fácil/divertida. Queremos saber por que ela tem as qualidades, usando o que temos discutido na disciplina).

2. Em seguida, vá nas visualizações de ao menos 2 colegas e poste (i) se você concorda com a desconstrução que ele fez da visualização, e (ii) um ponto positivo e um negativo diferentes dos que ele já mencionou.

Faça a postagem até a domingo 26/10, 23h59 e os comentários durante a semana seguinte.

Parte 2: perguntas sobre 10 mil das músicas no mundo

Músicas gravadas são um artefato cultural/artístico presente em muito do nosso dia. Seja no carro, na TV, em casa para relaxar, no trabalho para se concentrar ou na festa, há música ao seu redor. O tipo de música que alguém escuta geralmente tem bastante informação sobre a identidade dessa pessoa. Além disso, olhando para o que é produzido/consumido de música em um grupo ou época, é possível dizer bastante sobre a identidade desse grupo ou época.

Estima-se que haja por volta de 20M de músicas em formato digital. O ponto de partida de nosso 2o problema é uma amostra de destas músicas presente no [Million Song Dataset](#). A partir destes dados, queremos que você conceba perguntas e visualizações que examinam essas perguntas usando os dados.

Neste problema, não usaremos o MSD inteiro. Usaremos o Million Songs Subset, que é um subconjunto de 10K músicas mais fácil de processar. Ele existe em vários formatos, mas o que sugerimos é que você use o BD SQLite que está [neste link](#) (fizemos uma descrição das tabelas [aqui](#) e um código-exemplo pra acessar os dados [aqui](#)). Outros dois txts que podem ajudar e são simples de usar são o de todas as músicas [por ano](#) e o do [local de origem de cada artista](#) (mas lembre que não queremos mapas ainda). Dê uma passeada pelo site do MSD para tentar entender os campos nos dados, e poste no piazza suas dúvidas.

Considerando esses dados, conceba *5 perguntas* que você julga interessantes e que você vê que possam ser respondidas a partir dos dados. Lembre que você poderá usar tanto atributos que já estão prontos quanto outros derivados a partir destes.

Para cada pergunta:

- Justifique a pergunta dizendo a quem ela interessa
- Explique que atributos dos que já existem ou dos que podem ser calculados serão usados na visualização para responder a pergunta

Algumas sugestões:

- As perguntas não precisam envolver testes estatísticos ou modelos. Podem ser baseadas em identificações e sumarizações e comparações de valores/tendências/características de amostras que possam ser constatadas visualmente apenas.
- Queremos perguntas que envolvam mais de duas variáveis. Procure pensar se é possível categorizar parte dos dados. E agregá-los? Filtrar, talvez?
- Evite questões que seriam respondidas com mapas ou análises de redes/grafos. Lidaremos com esses tipos de análises no problema 3.

Na aula da segunda-feira 27/10, discutiremos as perguntas de todos. Suas perguntas devem estar prontas e documentadas em um txt no seu repositório **até a segunda-feira 27/10 23h59**.

Parte 3: visualizações em papel

Faça esboços usando lápis (ou caneta) e papel de visualizações que respondam a 3 das perguntas que você elencou. Fotografe ou escaneie os esboços e os adicione a seu repositório, documentando que pergunta está sendo respondida em cada esboço.

Comentários:

- Cada pergunta pode ser respondida por uma ou mais visualizações em conjunto, mas você deve produzir respostas separadas para as 3 perguntas.
- Tudo bem se você não for bom de desenho, mas procure tornar sua figura legível. Use canetas de mais de uma cor, por exemplo, e linhas de mais de uma espessura.
- Escreva na figura o que não estiver claro de ser visto de cara.
- Cuide no escaneamento para que a imagem não esteja tremida nem escura demais.
- Não faça o esboço no computador.

Discutiremos os esboços na aula da quarta-feira 29/10, e **as imagens devem estar postadas até a quinta-feira 30/10, 23h59.**

Parte 4: Código

Implemente um relatório usando RMarkdown + ggplot2 (+ quaisquer bibliotecas que você queira) que responda a 2 das perguntas que você criou na parte 2 implementando as visualizações que você criou na parte 3 deste problema. O relatório pode ser sucinto, mas deve, além de mencionar as perguntas e apresentar as visualizações, apresentar a sua resposta para a pergunta e a sua interpretação desta resposta. (Isso é, o que ela significa).