

Análise de Baseball

1. Links

Versão inicial	https://public.tableau.com/profile/m.rcio.souza.de.oliveira#!/vizhome/NDFDSI2-ProjetoFinalBaseball/Histria
Versão final	https://public.tableau.com/profile/m.rcio.souza.de.oliveira#!/vizhome/NDFDSI2-ProjetoFinalBaseball-REV_A/Histria

2. Resumo

O objetivo foi trabalhar com as características disponíveis dos diversos jogadores da liga americana de baseball, como altura, peso, estilo (canhoto, destro, ambos) e tentar encontrar correlações com os dados de rebatidas normais e home runs fornecidos de cada um deles.

As principais conclusões que consegui tirar da análise dos dados são que a grande maioria dos jogadores são destros e que, apesar de tudo, são os jogares canhotos que atingiram um grande índice de rebatidas e *home runs*. Os pesos e alturas dos jogadores da amostra estão correlacionados como eram de se esperar (quanto maior a altura, maior o peso) e jogadores mais baixos em geral (com exceção dos ambidestros) tendem a ter maior performance em rebatidas.

3. Design

Optei por iniciar a história com uma visualização do ranking dos jogadores, onde o critério seria o total de rebatidas (normais + *home runs*) para despertar a curiosidade do leitor a tentar entender porque determinado jogador está mais acima ou mais abaixo no ranking.

Depois resolvi destacar as principais distribuições dos dados, como o estilo dos jogadores, representados num gráfico de pizza para ilustrar a grande maioria dos destros. Já a distribuição dos pesos e alturas dos jogadores em geral escolhi um formato “*box plot*” para destacar os quartis e “*outliers*” e, finalmente, para a distribuição das rebatidas e home runs escolhi um formato de histograma, também destacando as principais estatísticas, para demonstrar que apenas poucos jogadores têm índices de rebatidas acima da média. Nesse painel é possível visualizar cada uma das distribuições filtradas por estilo, clicando no gráfico de pizza.

Em seguida, fiz algumas correlações entre os principais atributos com a performance em rebatidas e repeti o mesmo gráfico de pizza dos estilos para permitir a filtragem das visualizações. Com relação ao estilo, fiz a comparação da média de rebatidas com a média de *home runs* por estilo, usando um “*scatter plot*” para destacar a grande performance dos jogadores canhotos (essa visualização também pode ser usada para filtrar as demais). Também

plotei a relação entre a altura e rebatidas em linhas contínuas por estilo e mantive alguns “*box plots*” de pesos e alturas por estilo para posicionar o leitor de como está a distribuição desses dados enquanto avalia as demais visualizações sem precisar retornar ao painel anterior. Aqui também optei por manter o mesmo padrão de cores por estilo em cada visualização.

Finalmente, retorno para o ranking introduzido no início da história, agora que as principais análises foram feitas, para que o leitor possa olhar novamente essa informação com outro olhar e ratificar as conclusões realizadas. Também repeti as principais visualizações e estatísticas para ajudar nesse processo.

Após os *feedbacks* recebidos melhorei a apresentação da história criando um título principal e colocando sub-títulos em cada página para posicionar o leitor de como está organizada a história.

4. Feedback

Toda a coleta de *feedbacks* foi feita por meio do fórum da Udacity, e-mail e o canal do Slack durante o período de 1 semana e apenas 3 *feedbacks* foram recebidos. As perguntas sugeridas foram:

- O que você percebe na visualização?
- Quais perguntas você tem sobre os dados?
- Quais relacionamentos você percebe?
- Qual você acha que é a principal conclusão desta visualização?
- Há algo que você não entende no gráfico?

Seguem os *feedbacks* recebidos:

Feedback 1

- *O que você percebe na visualização?*

Várias estatísticas, está bem rico na parte quantitativa

- *Quais perguntas você tem sobre os dados?*

Qual o estilo possui os melhores jogadores?

- *Quais relacionamentos você percebe?*

Os ambidestros tem maior percentual de home-run e rebatidas.

- *Qual você acha que é a principal conclusão desta visualização?*

Existe uma altura ótima para ser um bom jogador

- Há algo que você não entende no gráfico?

Por que foram colocados esses gráficos boxplot de altura e peso. Acho que não é necessário explorar essa relação.

Feedback 2

O que gostei bastante na sua visualização foi a diversidade de gráficos e a plotagem apropriada de acordo com os dados.

Pontos que acho que podem melhorar: está faltando um título geral da história que indica o objeto e o objetivo central do trabalho e subtítulos nas seções. Digo isso por dois motivos: 1. por mais que tenha textos explicativos, muitas pessoas não gostam ou não tem tempo de ler (infelizmente, é importante ter isso em vista). Meu conselho é manter os textos, pois estão ótimos, mas inserir chamadas menores (pode ser em boxes captions ou subtítulos) o tema central de cada seção. Tipo: “Quais os principais atributos dos jogadores?”, “Os atributos influenciam na performance dos jogadores?” (só sugestões genéricas para explicar o que quero dizer).

Feedback 3 (após réplica ao primeiro feedback sobre os “*box plots*” e inclusão do segundo feedback)

Acho que seguindo a sugestão da [autora do feedback 2] pode manter, pois aí você justifica a presença dos gráficos

5. Recursos

Segue o link para o tópico no fórum da Udacity:

<https://discussions.udacity.com/t/coleta-de-feedback-projeto-de-baseball/654708>