

Identificando codificações

Vamos dar uma olhada em algumas das visualizações de dados e ver se podemos identificar as codificações que foram usadas nesses gráficos. O gráfico que você vê na tela no momento, ele foi projetado pelo Projeto Marshall. O Projeto Marshall é uma organização de notícias, uma organização de notícias sem fins lucrativos, e que, há um tempo atrás, fez uma história intitulada Crime em Contexto. E dentro dessa história, você tem um gráfico de linhas, um gráfico de linhas que mostra a taxa de crimes violentos em diferentes cidades entre 1975 e 2015. E então você tem a variação dessa taxa ao longo dos anos.

Antes de chegar à codificação, a propósito, gostaria de apontar uma característica que eu realmente gosto sobre este projeto, que é o menu. Tudo certo. O menu de navegação no topo. Você pode notar que o menu foi integrado, foi misturado com a introdução do gráfico em si. Você pode variar os diferentes elementos na introdução ao gráfico, e o gráfico vai mudar. Então, se você quiser ver, por exemplo, em vez de taxas, você quer ver as contagens totais, o número total de crimes violentos, você pode selecionar isso. E então a introdução ao gráfico vai mudar, e o gráfico vai variar em conformidade. E então se você quiser ver tipos específicos de crimes, por exemplo, se você quiser ver homicídios especificamente, você pode selecionar isso a partir do menu superior. E então todos os elementos no texto de introdução vão mudar em conformidade. Você também pode alterar a cidade, e se você quiser ver os números não de Milwaukee, mas, por exemplo, de Miami, onde eu moro, ou Miami-Dade County, então eu vou selecionar Miami-Dade County. Eu posso ver isso agora, neste momento, que a taxa de homicídios em Miami-Dade County diminuiu 53 por cento entre 1975 e 2015. E o intervalo de tempo também pode mudar. Enfim, esta é uma maneira muito interessante de apresentar os dados.

Mas de todo modo, vamos focar na codificação, então vamos nos concentrar no próprio gráfico. Que codificações, entre os que eu lhe mostrei antes, você pode identificar aqui? Vou apenas dar uma pista, ou uma ideia. Primeiro de tudo, perceba a própria linha, certo? A linha que representa a mudança, ou neste caso, homicídios em Miami-Dade de 1975 a 2015. A codificação nesse caso é a posição, mas por que posição? Bem, pense como um gráfico de linha é criado. Um gráfico de linha, a camada de andaime de um gráfico de linhas é geralmente um eixo vertical, que mede o que é, o que você quer medir, mais homicídios, menos homicídios, e, em seguida, o eixo horizontal é, neste caso, é tempo, certo? É ano a ano, certo? O que você faz em um gráfico de linhas normalmente, conceitualmente, é primeiro de tudo, colocar toneladas de pequenos pontos, cada um deles correspondente a um ano, sobre o eixo horizontal. Um ponto de 1975, 1976, 1977, e assim por diante. Você coloca os pontos no eixo horizontal, você muda a posição desses pontos de acordo com os anos. E então o que você faz é mudar a posição Y, a posição vertical ao longo desses pontos de acordo para a métrica que você quer medir, neste caso, as taxas de homicídio. Quanto maior for a posição de um ponto, maior é a taxa de homicídios nessa cidade nesse ano em particular.

Mas então o que fazemos é ligar esses pontos com linhas, mas a codificação ainda é a posição dos pontos que marcam os anos, certo? Como subproduto, embora, como um subproduto de ligar esses pontos, também poderíamos dizer que uma codificação

secundária de um gráfico como este é a inclinação, a inclinação da reta, o ângulo das linhas é igualmente uma pista de como, você sabe, interpretar ou ver a variação das taxas de homicídio ano a ano. Mas a codificação primária ainda é a posição, certo? A posição destes pontos identifica a taxa de homicídios em cada ano específico. Poderíamos também dizer, aliás, que neste gráfico a tonalidade da cor é usada como uma codificação porque a tonalidade de cor, neste caso, vermelho e cinza, a tonalidade de cor é utilizada para identificar a cidade que você está destacando, neste caso Miami-Dade e as outras cidades. Todas as outras linhas cinzas que você está tendo, embora, no fundo, todas essas linhas correspondem a todas as outras cidades do conjunto de dados. Elas ainda são visíveis porque o designer destes gráficos queria colocar a linha que você selecionou em contextos. O que lhe permite comparar a linha que você está selecionando com todas as outras linhas no conjunto de dados. Mas a tonalidade da cor foi utilizada para identificar, mais uma vez, a linha em que você está interessado e todas as outras linhas que devem ficar em segundo plano.

Outro gráfico ou outra visualização de dados que podíamos usar para testar a nossa capacidade de identificar, para identificar codificações é essa produzida pela NPR, National Public Radio. Portanto, este é um gráfico que mostra um mapa que mostra, como o título diz, um aumento dramático na cobertura de seguro de saúde sob o ACA. O ACA é o Affordable Care Act, também conhecido como Obamacare. Dê uma olhada no mapa que você tem lá. Esse tipo de mapa, porque você está interessado em termos, os termos que usamos na visualização de dados, é chamado de coroplético, mapa coroplético. Mapas coropléticos usam, bem, meio que um alerta de spoiler, a codificação deste mapa é tonalidade de cor, certo? Quanto mais escura a cor, maior o número que essa cor representa. Assim, no mapa que você tem na tela no momento, a principal codificação deste mapa são tonalidades de cores, certo? Como você pode ver, os condados no mapa, os que têm cores mais escuras são os condados que têm maiores taxas de falta de seguro de saúde. Quanto mais escura, mais escura a cor, mais as pessoas daquele município em particular não tinham seguro de saúde em 2015. E quanto mais clara a cor, menor o número de pessoas que não tinham seguro de saúde em 2015. Além disso, este é um gráfico interativo. Então você pode saber, você pode passar a seta por cima e você também pode usar o controle de tempo para voltar no tempo e comparar as cores atuais, os tons atuais de cores de com a aparência do mapa em 2010. E você vai ver que mais municípios têm taxas muito mais elevadas de pessoas sem seguro, certo? Antes de o Affordable Care Act ter sido aprovado. Tudo certo. Então é tonalidade de cor, certo? Tonalidade de cor é a principal codificação aqui.

Mas há uma outra, porque há um outro gráfico na tela no momento. Dê uma olhada neste tipo de gráfico de barras que você tem bem debaixo da cor, a legenda de cor. Isto não é realmente um gráfico de barras, embora ele usa barras para representar os dados, isso é chamado um histograma. Um histograma é um tipo de gráfico de barras, uma variação do gráfico de barras que é utilizado normalmente para representar a distribuição de uma variável. O número de contagens, que, neste caso, o número de condados que você tem dentro de cada faixa específica de uma taxa de seguro, certo? Então, basicamente o que este quadro está lhe dizendo, se você prestar atenção nele e considerando que a altura é método de codificação aqui, a altura destas barras é proporcional à contagem do número de municípios que você tem dentro destes tipos de "beans" de dados. Isto está basicamente dizendo que em 2015 existiam 110 condados que tiveram uma taxa de falta de seguro de saúde entre 0 e 5 por cento, mais de 1.000 condados que tinham uma taxa

entre 5 por cento e 10 por cento, cerca de 1.000 condados que têm 10 por cento a 15 por cento, 500 condados ou algo assim entre 15 e 20, 182 condados entre 20 por cento e 25. E 48 condados que têm uma taxa de mais de 25 por cento das pessoas sem seguro de saúde. Dê uma olhada na aparência deste histograma. Se voltarmos no tempo para 2010, você vai notar que as barras na extremidade direita do espectro se tornam muito maiores, tornam-se muito mais altas, mostrando que há mais condados com maiores taxas de falta de seguro de saúde. Mais uma vez, o método de codificação neste gráfico particular, este histograma, é a altura, a altura das barras é proporcional à contagem.