

Por que o jornalismo de dados não é mágica

Neste primeiro módulo você vai aprender como encontrar dados e como baixá-los. E nas próximas semanas você vai aprender a limpar esses dados, como explorá-los, como transformá-los, como visualizá-los, como contar histórias baseadas neles.

É bastante trabalho, ou pelo menos parece ser bastante trabalho, e uma das minhas experiências como professor de jornalismo de dados e visualização de dados é que muitos alunos, na primeira vez que eles encaram e começar a aprender sobre jornalismo de dados, se sentem um pouco sobrecarregados pela quantidade de habilidades necessárias para produzir jornalismo de dados e a quantidade de conhecimento que é necessário para isso: jornalismo de dados envolve aprender sobre estatística, aprender programação às vezes, aprender as ferramentas, aprender conceitos e princípios, aprender como lidar com as fontes, como consultar especialistas, como falar com profissionais de estatística e cientistas de dados que sabem muito mais do que nós sobre os dados com os quais lidamos todos os dias.

Mas meu objetivo aqui é basicamente falar para você que começar não é tão difícil. Para te deixar animado sobre jornalismo de dados e visualização de dados, deixe-me mostrar um exemplo de como é fácil simplesmente começar.

Em 2006, Hans Rosling, professor de Saúde Internacional no Instituto Karolinska, na Suécia, deu uma palestra famosa nas TED Talks, na conferência TED, apresentando dados sobre saúde global. Ele comparou, por exemplo, a expectativa de vida em diferentes países ao redor do mundo com as taxas de mortalidade e as taxas de mortalidade infantil. Ele falou sobre vários padrões na demografia em todo o mundo, e nesta palestra ele mostrou ao público várias belas visualizações interativas e animadas. Meu objetivo nesta parte da aula é mostrar a você como é fácil hoje em dia produzir visualizações tais como as que Rosling usou naquela apresentação.

O site do professor Hans Rosling é o gapminder.org. Na seção de vídeos do site você pode ver todas as palestras que ele deu ao longo dos anos e você também pode baixar alguns dos dados que ele usou em suas apresentações. Para esta pequena demonstração eu de fato baixei um conjunto de dados a partir deste site que reproduz ou é muito semelhante ao conjunto de dados que ele usou em sua famosa apresentação em 2006. Então, neste conjunto de dados que você pode baixar no site do curso, temos uma coluna com os nomes dos países, outra coluna com os nomes dos continentes, outra coluna com os anos. E então nós temos as taxas de fertilidade, a expectativa de vida e população de cada um desses países.

A ferramenta que eu vou usar para criar uma visualização que é muito semelhante à que Rosling usou em sua famosa apresentação, eu vou usar Flourish, uma das ferramentas que você vai aprender ao longo deste curso. Se você quiser usar Flourish, você não precisa aprender e acompanhar por este vídeo, esta é apenas uma demonstração, mas se você quiser começar, basta acessar a ferramenta e você estará pronto para começar, a ferramenta é gratuita.

Eu já fiz isso, eu já criei um perfil no Flourish. Então, eu venho aqui em "new visualization" (nova visualização), porque quero criar minha própria visualização como a de Hans Rosling sobre saúde global. Eu vou criar um gráfico de dispersão, que é um tipo básico de visualização de dados. Eu vou clicar no gráfico de dispersão e depois de fazer isso eu vou a "Data" (dados), eu vou importar os meus dados e vou encontrar o meu conjunto de dados que está nomeado como GapminderDATA_MOOC.

Basta fazer o upload dos dados. Uma vez que eu faço isso, eu importo publicamente, eu começo a selecionar os dados que eu vou visualizar. Então, primeiro de tudo, eu preciso dizer ao Flourish o que eu quero colocar no eixo X, no eixo horizontal. Então, no eixo horizontal eu vou colocar as taxas de fertilidade, país por país. Então tem o número da coluna, coluna D, então eu vou escrever aqui "D". No eixo Y, que é um eixo vertical, eu vou colocar expectativa de vida, eu vou colocar coluna E. Nome, que será o nome de cada um dos países, será a coluna A. Cores, eu quero que esses pontos que vão aparecer aqui sejam coloridos por continente. Então eu vou colocar coluna B, que será a cor. O tamanho das bolhas, eu quero que o tamanho das bolhas seja proporcional à população de cada um destes países, então isso será a coluna F. E, em seguida, a forma, eu não vou mudar isso. Eu vou rolar para baixo aqui e eu vou ao campo que diz "Time" (tempo). A coluna do tempo no meu conjunto de dados é a coluna do ano, que é a coluna C, então eu vou clicar em C. Depois de eu clicar em "return" (voltar), quando eu voltar para a pré-visualização, eu preciso fazer um ajuste aqui para fazer as bolhas ficarem muito, muito grandes, certo?

Não se preocupe se você não conseguir acompanhar, você vai aprender esta ferramenta mais tarde, no decorrer do curso. Mas o resultado é basicamente uma visualização que é quase idêntica àquela que Hans Rosling usou em sua famosa apresentação.

Aqui você tem o mundo em 1952, cada uma dessas bolhas corresponde a um país. O eixo horizontal é o número de filhos por mulher, que representa a taxa de fertilidade. E o eixo vertical é a expectativa de vida em anos, a expectativa de vida ao nascer. E aqui você tem um controle deslizante de tempo. Você pode rolar para a direita e dar uma olhada no mundo em 2012. Observe a diferença. Em 1952, mais países estão aqui, com muitos filhos por mulher, altas taxas de fertilidade, expectativa de vida baixa. Mas quando vamos para 2012, muito perto do presente, você vai notar que a maioria dos países está concentrada no canto superior esquerdo, que significa expectativa de vida alta e taxas de fertilidade muito baixas.

Este é o poder do jornalismo de dados, este é o poder da visualização de dados, e também o quão fácil é começar neste mundo.