Why data journalism isn't magic

In this first module you're going to learn how to find data and how to download it. And then in future weeks you will learn how to clean up this data, how to explore it, how to transform it, how to visualize it, how to tell stories based on it. It's a lot of work, or at least it sounds like a lot of work. And one of my experiences as a professor of data journalism and data visualization is that many students the first time that they face and start learning about data journalism they feel a little bit overwhelmed by the amount of skills that are required to produce data journalism and the amount of expertise that is required. Data journalism, learning about statistics, learning about coding sometimes, learning about tools, learning about concepts and principles, learning about how to deal with sources, how to consult with experts, how to talk to statisticians and data scientists who know much more than we do about the data that we handle every day.

But my goal here is to basically tell you that getting started is not that hard. To get you excited about data journalism and data visualization let me show you an example of how easy it is to just get started. In 2006, Hans Rosling, a professor of International Health at the Karolinska Institute in Sweden, gave a famous talk at the TED talks at the TED conference. Presenting data about global health. He compared for example life expectancies in different countries all over the world to mortality rates and infant mortality rates. He discussed several patterns in demographics all over the world, and in this talk he showed the public tons of beautiful interactive and animated visualizations.

My goal in this part of the lecture is to show you how easy it is nowadays to produce visualizations such as the ones that Rosling used in that presentation. Professor Hans Rosling website is gapminder.org. In the video section of this website you can see all the lectures that he gave throughout the years and you can also download some of the data that he used in his presentations. For this little demonstration I actually downloaded a data set from this website that reproduces, or is very similar to the data set that he used in his famous 2006 presentation. So in this data set which you can download from the course's website, we have one column with the country names, another column with the continent names, another column with the years. And then we have fertility rates, life expectancies, and population of each one of these countries.

So the tool that I'm going to use to create a visualization that is very similar to the one that Rosling used in his famous presentation, I am going to use Flourish one of the tools that you're going to learn throughout this course. If you want to use Flourish, you don't need to learn you don't need to follow along this video, this is just a demonstration but if you want to get started you can just sign in for the tool and you will be ready to go, the tool is free. So I already did that, I already created a profile in Flourish. So, I'm going to go to "new visualization" because I want to create my own Hans Rosling visualization about global health. I'm going to create a scatter plot which is a basic kind of data visualization, I'm going to click on the scatter plot and after I have done this I'm going to, going to go to "Data", I'm going to import my data and I'm going to find my data set which is called GapminderDATA_MOOC. Just upload that, once I do that I import publicly, I start selecting the data that I'm going to visualize.

So first of all, I need to tell Flourish what do I want to put, what I want to put on the x axis, on the horizontal axis. So on the horizontal axis I'm going to put fertility rates country by country. So there will be column number, column D, so I'm going to type here "D". On the y

axis, which is a vertical axis, I'm going to put life expectancy, I'm going to put column E. Name, that will be the name of each one of the countries that will be column A. Color, I want these dots that are going to be appearing down here to be colored by continent. So I'm going to put column B, that will be the color. The size of the bubbles, I want to be the size of the bubbles to be proportional to the population of each of these countries, so that will be column F. And then the shape, I'm not going to change that. I'm going to scroll down here and I'm going to go to the, a field that says "Time". The time column on my data set is the year column which is column C. I'm going to click on C. After I have clicked on return, when I go back to preview I need to do an adjustment over here to make the bubbles really, really big. Right.

Don't worry if you can follow along, you will learn about this tool later on throughout this course. But the result is basically a visualization that is almost identical to the one that Hans Rosling used in his famous presentation. Here you have the world in 1952, each one of these bubbles corresponds to a country, the horizontal axis is the number of children per woman, that is the fertility rate. And then the vertical axis is the life expectancy in years, the life expectancy at birth. And then over here you have a time slider. You can scroll to the right and you can take a look at the World in 2012. Notice the difference. 1952 most countries are over here, tons of children per woman. High fertility rates, low life expectancies. But then when we move to 2012, very very close to the present, you will notice that most countries are closer on the upper left corner meaning very high life expectancies and very low fertility rates. This is the power of data journalism, this is the power of data visualization, and this is also how easy it is to get started in this world.