

FIVE TIPS FOR JOURNALISTS
DOING DATA
VISUALIZATION

It's not just for the designers and 'Web folks'

BY **STEPHANIE YAMKOVENKO**

General Electric is doing it. Statistical software behemoth SAS is selling software for it. Harvard is teaching courses about it. Data visualization seems to be everywhere.

Journalists need to get involved to provide something that can easily go missing in data visualization: the story. Data visualization should not only describe a dataset, it should also provide context and depth to the data. It should tell a story.

I will share my five tips on how to tell compelling stories with data visualization. Although I wouldn't consider myself an expert on the topic, I recently won The Economist-Nielsen Data Visualization Challenge. (See more on the challenge and winning entry at innocentive.com/NielsenFinalists.)



Stephanie and Bogdan Yamkovenko made this presentation for The Economist's Nielsen Data Visualization Challenge. The examples shown throughout this article are from their winning entry that correlated the level of a country's consumer confidence with its press freedom index.

HOW FREEDOM OF THE PRESS AFFECTS CONSUMER CONFIDENCE

DATA: Consumer Confidence Index (CCI) is a convenient way to understand the perceptions of individuals about the health of their economy. Nielsen Global Survey provides the CCI and information about other consumer behavior including plans to purchase products and use social media to inform their purchases.

QUESTION: In looking at the global perceptions, we noticed that certain countries (such as Saudi Arabia, China and UAE) **seemed overly confident in their economy**. Why were they so confident? What do they have in common? We had an idea that perhaps the information these citizens receive from their press was shaping their confidence.

APPROACH: We supplemented the data set with the Press Freedom Index from Reporters without Borders (as well as unemployment data), used multivariate data analysis to analyze it and found the following...

Bogdan Yamkovenko, PhD & Stephanie Yamkovenko, MMC

WHAT IS THE EFFECT OF INTERNET?

Before the Internet, it was very easy to control the flow of information. How does the Internet affect confidence, free speech and purchases?



People want **truthful, honest** information and if they cannot get it from their country's press, they will turn to the Internet and social media to find it.

* Correlation between Internet penetration and Press Freedom Index $r = -0.615$ ($p < 0.001$).

** Correlation between Press Freedom Index and likelihood of making a purchased based on social media website $r = +0.796$ ($p < 0.001$).



The Internet, social media and a free press give us the information we need so that we can know the **truth** about our economies.

"INFORMATION IS THE CURRENCY OF DEMOCRACY"
-THOMAS JEFFERSON

Sources: 1) Reporters without Borders. 2011-2012 World Press Freedom Index. www.rsf.org. 2) Nielsen Global Survey. 1st Quarter, 2012. 3) The Economist. Output, prices and jobs. www.economist.com.

The challenge focused on analyzing a dataset provided by Nielsen and telling a story using data visualization. After seeing the challenge advertised in The Economist, I decided to partner with my husband, Bogdan Yamkovenko, a researcher and assistant professor with an affinity for statistics.

Having never collaborated on a professional project before (and the fact that neither of us currently works as a graphic designer), we weren't too confident about our prospects for winning. When we got the news that we won and had the opportunity to present our findings at The Economist's World in 2013 Conference in December, I realized there is an important role for journalists in data visualization. Using my experience with the challenge, I developed the following tips for journalists.

1 DON'T IMAGINE AN INFOGRAPHIC

The first tip is to stop thinking about an infographic. It can be difficult, because ultimately graphics will be a big part of data visualization, but for the time being focus on the data. It's similar to not writing an article before interviewing sources and doing research.

Bogdan and I began our work on the challenge with a brainstorm about the Nielsen global dataset, which consisted of the consumer confidence index and other data about consumer spending and purchasing habits. Look at the data and see what is there and whether you can see any potential stories that need telling.

2 WHAT'S NOT IN THE DATASET?

Chances are the dataset itself is not complete. Can you supplement the dataset with other data? Think of ways that supplemental data could offer context to the current dataset. Again, journalists rarely write an article after interviewing only one source. Why would that be different with data?

We decided to supplement the Nielsen dataset with the widely available economic indicators such as unemployment rates and the current account balance as a percentage of GDP, which is information on the final pages of every Economist issue. By adding this information, we were able to look at whether the consumers' confidence in the economy matched up with the country's actual economic performance.

3 WHAT IS GOING ON WITH THE OUTLIERS?

Take a look at the dataset that you have supplemented with other relevant data and see whether you find any patterns, outliers or strange con-

nections. After supplementing the Nielsen dataset, we started looking at the outliers (the countries with the highest and lowest consumer confidence). We noticed that countries that had high confidence in the economy were not necessarily the best performing economies. Countries such as Saudi Arabia and Egypt had high confidence, but their economies weren't doing that great.

At this point, we revisited tip 2. We wondered whether democracy was playing a role in the citizens' confidence. We decided to include the Reporters Without Borders Press Freedom Index in our analysis.

4 USE STATISTICS

Yes, many of us went into journalism because we don't like numbers. In graduate school, one of my journalism professors assigned us the textbook "Statistics for People Who (Think They) Hate Statistics." Even if numbers are not your forte, it helps to have

a basic understanding of descriptive and inferential statistics; I still revisit that textbook for a refresher on statistics.

Whether using Excel or SPSS — or, for the real nerds, R — to analyze the data, take

that countries with the highest consumer confidence also have the most restricted press. Bogdan was able to do much more advanced statistical analyses using multilevel modeling and other techniques, but that

I've thought a lot about data visualization since winning the challenge, and one thing I feel strongly about is the important role for journalists. With every story we tell, we are using the skills necessary for data visualization: gathering information (data) from multiple sources, examining the facts, arranging it coherently, and providing readers with context and depth that will help them understand the story.

the data, find correlations and run regressions. This step may require outside help, but it is essential to finding meaning in a dataset that goes beyond reporting numbers and descriptive statistics.

With just a simple correlation, we found

correlation was the finding that gave us a compelling story to tell.

5 TELL THE STORY

This should be the easy part for journalists. In fact, at this point an infographic-

CONFIDENCE + PRESS FREEDOM

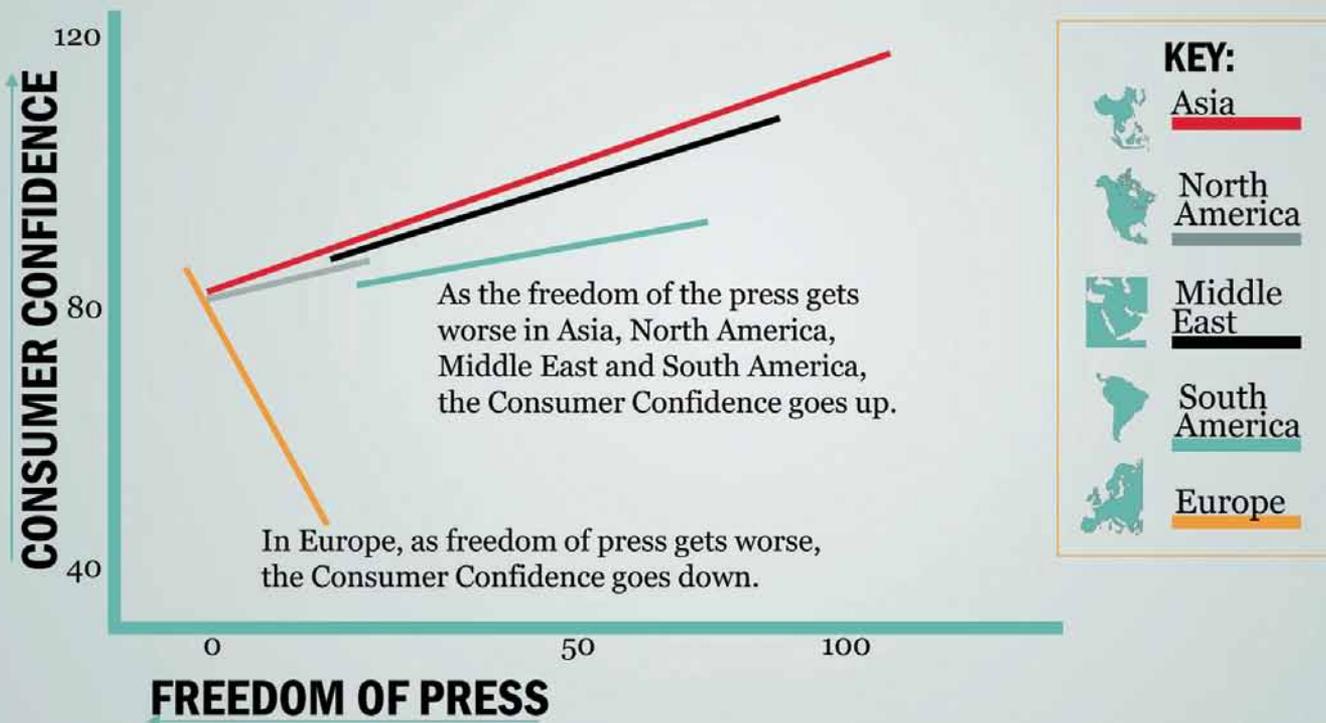
When the government **controls** the information flow, it can report as rosy a picture of the **economy** as it likes. This is why we see the following*:



*Correlation between CCI and Press Freedom Index is $r=0.465$ ($p<0.01$).

REGIONAL INFLUENCE ON CONFIDENCE

Digging deeper, when predicting Consumer Confidence Index using the Press Freedom Index by region, a clever multilevel modeling technique allows each region to be modeled separately but in one model.* This reveals an unexpected relationship.



*This unusual finding could be the result of a limited sample (N=34). The length of the lines represents different sample sizes within regions.

ic should still not be a concern. Tell the story you found in the dataset by writing it (either in story form or bullet points). Once you have the story, figure out how graphics, illustrations and charts can enhance the story and help readers visualize the findings.

My first job out of graduate school was working as a graphic designer, so I was able to use those skills to design our winning submission. If you don't have that background, work with a graphic artist or an in-house designer to develop the visuals that will enhance your story. Our submission to the challenge was a PDF with several "slides" that presented our story. The story will dictate how you will present the findings, whether it's an infographic, an article with charts and illustrations, an animated graph, etc.

THE ROLE OF JOURNALISTS IN DATA VISUALIZATION

I've thought a lot about data visualization since winning the challenge, and one thing I feel strongly about is the important role for journalists. With every story we tell, we are using the skills necessary for data visualization: gathering information (data) from multiple sources, examining the facts, arranging it coherently, and providing readers with context and depth that will help them understand the story.

Don't let a fear of numbers or graphic design get in the way. Get some help, collaborate or work on some new skills. With so much data out there, imagine the stories you can be telling and visualizing. 💎

Stephanie Yamkovenko is a journalist focusing on health care practice and policy in the Wash-

ington, D.C., area. Interact with her on Twitter: @S_Yamkovenko.

GET DATA VIZ HELP

New to data visualization?

No problem. It's not just a skill for "the digital folks" or designers. All journalists can benefit from telling visual stories with data.

SPJ will present its first free webinar on data visualization for beginners in February. Register at spj.org. Hurry — space is limited.