

14th Annual International Symposium on Online Journalism

Day 2, April 20, 2013: Afternoon Session - 2:30-4:00 p.m. *Data Visualization: Creating a New Language to Communicate Big Data*

Chair: Alberto Cairo, Professor of Practice, **University of Miami**

Research Panelists:

- **Hannah Fairfield**, Senior Graphics Editor, **The New York Times**
 - **Chiqui Esteban**, Visual Journalism, **Boston Globe**
 - **Scott Klein**, Editor of News Applications, **ProPublica**
 - **Kim Rees**, Partner & Head of Data Visualization, **Periscope**
 - **Lisa Strausfeld**, Global Head of Data Visualization, **Bloomberg**
-

Hannah Fairfield: Thank you. I wanted to talk about immersive storytelling as well as information graphics and data visualization, because I was able to be part of a really special project last fall that really worked on taking a lot of the wonderful resources in an entire newsroom and combining them with some pretty special information graphics and being able to put together a really interesting project that many of you may have seen. It was called Snow Fall. And it came out in December. And we were pretty excited about it. We weren't quite aware of the splash that it actually made in journalism until after it actually came out, but it was a wonderful project to work on. And the information graphics that I worked on and the sort of immersive factor of them is something I want to talk about today.

So, one of the things that when we started thinking about what we might want to be able to make, you know, you kind of look for inspiration from your past, things you've read, things you've seen, people you've talked to, you know, people in the room. You get kind of excited about things like that. And we knew we had an interesting project, and we knew that we wanted to make something really immersive. And I started thinking about some of the things that I had looked at in the past couple of years. And one of them was this book called *The Invention of Hugo Cabret*. And I don't know if any of you have seen it, but it's really interesting. And it's a book that also has charcoal sketches in it. So, the storyline is actually told through charcoal sketches and words. So, you're reading along and you see two or three or four pages of charcoal sketches, and then you turn the page, and all of a sudden, it's words. And then after that, it's charcoal sketches again. And it's wonderfully immersive. And the pictures carry the story. So, what we really wanted to do is to be able to have these graphics really carry the story. It was really important to me.

So, the project had kind of a modest beginning. In the very first meeting, we actually came up with a sketch like this. This was done on the yellow pad of

14th Annual International Symposium on Online Journalism

Steve Duenes, who's the graphics director at The New York Times. And we sat together. There was just a handful of us. And we thought, what do we want to do with this? What do we want to make it? And we were all very excited from the beginning to be able to use all the elements of the newsroom and put them altogether in one piece. And we started to sketch out what that might be and talk about what that could look like. And this was another early sketch with some dummy text and some beautiful scene setting photography. And we thought, ultimately we're going to have some information graphics in there. And those are in the green. The green boxes were going to be the information graphics.

And so once we started with this, we knew that we wanted to have those infographics, again, like I said, really carry the story. And one of the most important things that I thought the first infographic had to do is to take you to this mountain. The story is you're on a mountain with a lot of skiers, and you know an avalanche is coming because we told you that from the beginning, but we want to be able to let you know what it feels like to be in that mountain. So, maps do that extraordinarily well. And we ended up wanting to set the scene with a map that looked like this.

And Jeremy White, who's a photographer at The New York Times, he had an idea that he really wanted to be able to bring you to the mountain. Not just show you a picture of it. Not just show you a map of it. But actually bring you there as if you were coming in from — driving through it or flying over it. And he's an exceptional cartographer, and he looked around at the resources that he had, and he ended up making an actual model of it using digital elevation. And this is what he started with. So, this is the base layer. This is data. This is your data visualization. It's taken from LIDAR data and it's just a raw model. So, it doesn't really look like a mountain yet. So, in order to make it look like a mountain, he looked around for satellite imagery. And we ended up putting up a satellite orthoimagery on top of it. The best satellite imagery that we had was taken in the summer, and the actual avalanche had happened in the wintertime. So, in order to correct for that, Jeremy actually put on the snow. So, as you watch it sort of immersively, you can really get a sense of what it looked like. And it feels like you're in an airplane flying through it, but it's a digital elevation model that he's actually done.

So, looking at that, we wanted it to feel like the day that the actual avalanche happened. So, Jeremy added some fog and some haze to make it feel like the day. And then, one of the things that I knew was really incredibly important with this was being able to understand that they are at a ski resort and they are focused on the ski resort, but the group of people that were involved in the avalanche went out of bounds. They were out of bounds of the ski resort. So, you want to define the boundaries. You want people — you want your readers to be able to understand what they're seeing. So, we put a yellow line around that defined the boundary, and then we took you over the edge, immediately over the edge. You're right on the top part of the mountain. And then we swing you around and we identify exactly where it

14th Annual International Symposium on Online Journalism

happened, so there you are, top of Cowboy Mountain at Tunnel Creek. We've set the scene for you.

So, that was our first data visualization. It didn't feel like big data. It wasn't dots on a map. It wasn't pie charts. It wasn't things like that. But there's a lot of data behind that. There's a lot of thought. There's a lot of editing behind that.

So, the next one that we focused on a lot, that for me the transition was harder to think about, we spent a lot of time thinking about how to actually build this. And that was the skiers paths. John Branch, who wrote the piece, had spent a lot of time with the survivors talking about what exactly happened at the top and who was where. And so when I first read the draft of the story, keeping all those people separate and understanding who was where when the avalanche happened, it was very difficult for me, and I wanted to make sure that no reader ever had to have that kind of difficulty. And I knew an information graphic would be able to provide that assistance, that help. So, what we ended up doing is we brought you back to the top of the mountain. Which you remember from the last information graphic, you're standing right there. You hike in. but the important thing about this is that instead of being totally immersive, like the other one, we need you to keep reading the story. The story is you're scrolling down, and as you're scrolling, you're getting information. You're getting information about the skiers.

So, a small group of skiers ended up skiing away from the other big group. They didn't even know the avalanche had happened. They were wearing cameras on their helmets, and at one point halfway down the mountain, they stopped. And they looked around and they thought, "Wow, it looks like some snow is sliding. It really doesn't look safe back here." And we had that footage of them. So, that video was embedded right in there. You can watch it right here as you actually can follow their actual paths through the information graphic down.

Another skier who was actually caught inside the avalanche, she was also wearing a helmet, and so we have the footage of how she was skiing up to the point in which she stopped right before the avalanche happened. So again, that video is embedded in there. You can see her yellow line. She's right there. You're with Elyse skiing down the mountain. And then after Elyse, there are a couple of people. And one of the things when I first read the story that again made me stop was that there were some people who were there when the avalanche happened. They watched the avalanche, but it didn't come through the area where they were. But it seemed to me like all the skiers were kind of in the same place, so I didn't really understand that. But looking at this, you can see one path juts over to the side. There's a bunch of people who are together, and then there's a bunch of people — there's one path that actually goes over to the side in yellow right now, and those are the people who were up there when the avalanche happened, but were not caught in it, but they saw it go past.

14th Annual International Symposium on Online Journalism

So now, we've set the scene for the avalanche. You know where everybody is positioned on the mountain. John Branch has been building the anticipation, the drama of the story. And then our next big challenge, I really felt, was to have readers understand what it was like to be inside an avalanche; what that would feel like. And we struggled with it in a lot of ways. I read a lot on fluid dynamics and how avalanches flow and work, and we did a lot of particle simulations trying to figure out, how can we construe that to readers? The turbulence of being inside of it. What is it like? And ultimately, we were very fortunate in that John Branch put us in touch with a bunch of really amazing scientists in Switzerland. And what they ended up doing is telling us that they could potentially do a model for us of the actual avalanche. And they said, "We'd really like to do it for you, but what we really need is the actual elevation, an elevation model of the mountain, in order for it to be really accurate." And Jeremy White and I said, "Hey, we just happen to have one of those." And we sent it over, and in less time than you can possibly imagine, they sent us this beautiful thing, which is their output from their extraordinary avalanche simulation model. They gave us a bunch of these. They gave us real-time speed, velocity, height. It was so exciting with this dropped into our inbox. We just couldn't believe it.

And so, what we started to do with this is to think about, how can we then bring you into the avalanche? And we decided instead of putting the readers inside of it, we let you see what it was like to be right in front of it. So, we brought you back to the top of the mountain, and then we positioned exactly where the skiers were. So, you can see there's a clump and then there's two on one side. You can see where the avalanche actually came through. So, that's in real time. We're showing you in real time what it was like to be on top of that mountain. And then we have real data—48 miles per hour. We know exactly how much it weighed. It was amazing to have access to that kind of data that we could use to analyze this. And then 43 seconds later, you're at the debris pile. And you can watch it online. You see the whole thing go through. And then slowly, slowly, the debris pile ends up piling up. And that kind of, you know, before the recovery, that's really the end of the story. It's a very quiet moment.

So, what a lot of people have been asking me is, what's next? What does this mean? Does it mean anything? Does it mean everything? What's really happening? And my answer to that is that, there's really no better time, I think, to be in journalism. I mean, we have all these amazing tools at our fingertips, and the world is really our playground. And when you can build something fun like this, ideas spread like wildfire. And just I feel like it was a couple of weeks after we ended up launching this, the Washington Post came up with this wonderful, wonderful package on what cycling is like post-Armstrong. And they did an amazing job. And it doesn't look anything like Snow Fall, but, you know, perhaps it was clearly influenced by [it]. They learned from our mistakes and what we did, and they expanded it. It's really been wonderful. And you see other examples as well. The Atlantic Wire

14th Annual International Symposium on Online Journalism

teased the White House a little bit for coming out with something that looked [like] what they called the Snow Fall Effect. And then just actually this week, Microsoft came out with an interesting multimedia package in which they sort of used their own atmospheric mood-setting photography, which is pretty much what a newsroom looks like and probably a lot what the Microsoft offices look like.

So, like I said, journalism is — it truly is, I think, the best time in the world to be able to do exactly what we do in the way that we love to do it and to share all these ideas and build incredible stuff together.

[Applause.]

Scott Klein: Hi, everybody. (Put my water down.) As Alberto said, I'm Scott Klein. I work at ProPublica in New York. And just to repeat, I run a team of nine software developer/journalists who build what we call news applications. News applications are large interactive database projects that tell a journalistic story using software development instead of words and pictures. Our work combines software development with traditional reporting and journalistic data gathering, data analysis, like data journalists do, as well as data presentation, user-interface design, visual design.

Like all visual journalists like Alberto and Hannah and Kim, we ask ourselves a lot of questions when we are working on a project. We ask ourselves, what is our story? What is our data? What story does our data tell? How can we use interaction design to help our readers tell this story to themselves? This is something that Alberto pointed out.

Our work is primarily investigative in nature. So often, we are covering things that people don't necessarily know very well or are not very familiar with. So, we need to kind of find ways when we're working on a project to build interfaces that help people navigate through very complex data without either discarding data and only showing them summaries, and also in a way that makes them feel empowered, and empowered to explore, and not make them feel small and stupid.

One of the ways we do this is by taking advantage of levels of abstraction. And so we slice the data up into the general all the way down to the specific, and we hope to help users propel their way using interaction design, using visual design to construct a narrative for themselves that takes them from the general to the specific.

I wanted to show you a movie, but it would have taken all of my time, so I've kind of done a quick abstract of this movie. This is a movie or it will be slides from a movie by Charles and Ray Eames, the famous furniture designers, from 1977, called Powers of Ten. Powers of Ten starts from a part in Chicago, from a couple having a picnic, and zooms all the way out into outer space. This will go on and on. And what they invited you to notice in this

14th Annual International Symposium on Online Journalism

movie was how things changed as you zoom out. So, detail gives way to context. Information changes. The utility of the information changes. The nature of the information changes even though technically and quite impossibly you're always looking at the same thing. This will keep going on and on and on. So, it's a terrific video. It's on YouTube. They also zoom all the way out and then they zoom all the way back in to an atom and then beyond.

So, two levels of abstraction we talk about when we were designing a project at ProPublica is the far and the near. These are the extremes of the levels of abstraction for an application that we're building. The far is typically the landing page of the application. The one that tells you the big national story. Why am I on this page? What's the story I'm supposed to understand? What journalists might call the nut graph, the lead. These are on the landing page of the story. And the near is essentially the terminal page. The one that is closest to the bottom, the closest to you, the most specific data.

We, of course, stole this idea from cartography, which has a very different understanding of the words far and near; a much more reasonable understanding of the words far and near. But notice how when you're looking at a globe the information that you see, the actions that you might be able to take with that information are different than when you're looking at a U.S. map. A U.S. map where you might be able to understand the relative distances and sizes of the states and their capitals. Perhaps you can understand which states border the Great Lakes or how far trains have to travel or airplanes have to travel. Which is obviously different from a state map which has very different uses, very different information. A city map. A city map starts to be one where landmarks become very familiar to you. Your familiarity with the environment starts to increase the closer that you get down to the near. And this is also where you can start really putting utilities on top of the map. You can start overlaying, in this case, the bus map. So, I wouldn't want to put a bus map on a big national map of the U.S. You obviously couldn't put a bus map on a map of the globe. But when you are down to a city, you can start really doing interesting stuff in terms of news you can use.

And the zooms keep going further. So, this is my neighborhood in Brooklyn; a 19th century map of my neighborhood in Brooklyn. Well, my old neighborhood in Brooklyn. There's no laser or I could point out where I live. But a neighborhood map now is useful for knowing more about my neighbors, more about the context of the people who live in places I know very, very well, that I walk past every day. Where is the subway station? Where is the nearest grocery store? And I can even zoom these days, thank God, into people's homes. This should be familiar to anyone who's rented an apartment in New York City — the floor plan. This is my old apartment. I loved it. It was beautiful. I don't live here anymore. But the closer you get, you can get further and further and further, more and more into familiar territory, more and more into things with genuine utility.

14th Annual International Symposium on Online Journalism

Just two examples of maps with data overlays [that] I want to show you very quickly. This is a map of energy use in buildings in New York City that Columbia University made. And here on the city view, what can you see? Anybody? [Inaudible response.] It's all about Manhattan. Exactly right. And if you've ever been to New York, you know that Manhattan is where all the tall buildings are. So you can see that the researchers clearly have found that the more floors you have, the more you use energy. But the correlation becomes obvious when you're zoomed this far out and can see sort of the helicopter view of the city. I can zoom in. This is Park Slope. Again, this is my neighborhood. And I can see, again, you can see where the tall buildings are in Park Slope. But you can start to see gradations and energy use in different buildings. Pretty much the yellow and the orange are all brownstone buildings in Brooklyn. And as I zoom in further, I can now see things that are meaningful directly to me. So, my building is on this map. I can see my building's use of energy and what I might be able to do to change the energy use in my building, if I wanted to do that.

Another quick map. This is a 2009 electoral map for mayor of New York. And at this zoom, what can you see? You can see obvious correlations between class, ethnicity, race, and who voted for whom. And if you are from New York and you know these neighborhoods, these things probably are not a surprise to you ... or maybe they are. This may be a surprise to you. Here's Park Slope again, which happened to go Republican or whatever Bloomberg was then. [laughter] Sort of like a liberal stronghold, Park Slope went for Bloomberg. And you can even see it. This is a beautiful graphic that Matt Bloch and Archie Tse at The Times did. You can actually go down to my census block or my district and see that Bloomberg won by 50 votes, which is a lot in a block.

So the far and the near and levels of abstraction are—this is the best slide I could come up with for this—are also very familiar to journalists who talk when they're writing a story, a narrative story, in the inverted pyramid. So the inverted pyramid style, where you go from the general to the specific, from the very important facts down to the more specific facts. But news applications afford a new opportunity, kind of an Internet-only opportunity for journalists. Because when they talk about the inverted pyramid, a writer, he has to — a reporter has to think of a great example. He has to come up with a great anecdote for his story. He has to find one that will be meaningful to the most people. But he can only find one anecdote or maybe two really in his story. And if he's lucky, less than 1% of the people will — that anecdote will be meaningful to them. "Oh, yeah, I went to that high school," or, "I knew that person," or, "I'm also from Topeka, so this city story is meaningful to me."

New applications can tell thousands of stories. It can tell everybody's story. It can tell your story. You can find yourself. You're from Topeka, you can find Topeka. I'm from Brooklyn, I can find Brooklyn in a story. And our job is to

14th Annual International Symposium on Online Journalism

make it so that, again, through user-interface design, through interaction design, we make it so that you can find your story. You can still explore, but we encourage you to navigate through the app in a way that tells the journalistic story we're trying to tell.

This is an app that we built at ProPublica, my team as well as the car team, that used U.S. Department of Education data on educational opportunity. Educational opportunity as they define it is essentially access to math, a higher math, chemistry, physics, advanced placement, gifted and talented. A lot of programs that studies show lead to better outcomes. And the question that the app asks is — or the question that the hypothesis asked was, "To what extents are states and districts distributing educational opportunities fairly?" So, rich, poor, black, white, Hispanic. Are all kids getting equal access to these things or is there some predictor that shouldn't be there?

So, this is the landing page of the app. Think of this as the national map that you saw. This is the national map of our data. And you can see, in fact, on this that there is a list of states ... just like on a national map. And there are, I think, nine criteria. I think it started with seven. Nine criteria. It turns out that not every state or there are no perfectly good states and no perfectly bad states. There's a lot of gradations in between and some states were good at some things that other states were bad at. So, we made it so that you could navigate the different states. You could sort. If you're sorting by gifted and talented.... (Oops. Yeah, you gotta see that again.) Sort by gifted and talented, sort by AP enrollment. So, you can start understanding, again, that national map, which states are good, which states need to do some more work.

But the most important thing is this search box up at the top. We really want to encourage people to get down to levels of abstraction as soon as they understand what the nut of the story is, as soon as they understand that this is about educational opportunity and which schools are good at distributing, which districts are bad. We want them to find their own school, so we make it.... It kind of gets cut off, but it makes the sort of giant thing on the page.

So, I did not grow up in Santa Monica, but if I had grown up in Santa Monica, I would have gone to Santa Monica High. This is all sorts of stats about Santa Monica High. So, there's 26% of the students at Santa Monica High get free and reduced price lunch. 4% of the teachers are inexperienced. Maybe I didn't know those numbers, but I understand Santa Monica, so I now have an intuitive sense of what those numbers mean, because I can map it to my knowledge about Santa Monica High. We also.... If you hover over any of these bars, it will give you the state and district averages for these things, so I can see that Santa Monica High is about half of the state average for free and reduced priced lunch. Free and reduced price lunch happens to be the proxy that we used, and it's a very standard proxy to use to understand poverty. So, [if] a lot of kids take free and reduced price lunch, high poverty school. [If] none, it's a rich school.

14th Annual International Symposium on Online Journalism

So, there's one deeper level of abstraction. One way that you can navigate even beyond this. So, if this is the city map, maybe this is the bus map. So, if you click on 'compare to high and low poverty schools' on the left, we now show you two example schools in California, with Santa Monica High in the middle, Sequoia High on the left, Miramonte High on the right. (Got another minute. I'll be done in a second. Okay.) And on the left, and this tells our exact story. This is a prediction. It shows a relationship that shouldn't be there in an ideal case. As poverty goes up, access to educational programs goes down. And we also let you drive up the levels of abstraction as well and start to understand the ways in which the district has this relationship and the state as well.

So, that's all I got.

[Applause.]

Kim Rees: Thanks for inviting me. I want to dedicate my talk to Antonio Santiago, who you can see here. Last month, he was shot in the head and killed. He was 13 months old and he had just learned to walk. I'm not going to talk about periscope, but I do want to talk about one of the projects we recently did about gun violence in America. (And where's the...? Oh, here it is. Sorry. Didn't come prepared.)

The one thing I do want to talk about before anything else, before I talk about this project is, if I can tell you one thing today, it's make visualizations. Make visualizations. It's a tool that you can have in your tool box just like anything else you use, like video, like photography, anything else you use to tell your stories. Visualizations are.... They have this amazing ability to make people believe in what you're telling them. They help people join you on your journey of facts through your story. They can be very powerful. But I think that sometimes we get stymied by data. We get a little terrified of it. It can be kind of intimidating for people who don't work in data, who don't work with numbers. I think that you all are talking about, you know, statistics a lot and mathematics, and these things can be kind of scary if you've never used them before, but I don't think that they have to be that bad.

I think that we fall victim to this thing I call *death by disclaimer*. And that's when you get a load of data and you say, well, it's incomplete or it's not as robust as it should be. How precise should we get? How accurate do we need to be? And [you] start throwing around all these statistical terms until you become so overwhelmed that you just throw up your hands and give up and throw it all away. And I think that's a real tragedy and I want to challenge people to really start thinking about it in not such scary terms. I mean, I think that, you know, some of these presentations have ranged from very simple ways of exploring data to very complex ways. And we don't all have to start at The New York Times level of visualization. They make these

14th Annual International Symposium on Online Journalism

beautiful things, and they have wonderful resources to do those. But even at smaller levels, I think that we can make really understandable graphics and visualizations.

The piece that I want to talk about, I'm going to show a brief video of it, so I'll just let that play out and then I'll talk more about it.

[Video plays of "U.S. Gun Killings in 2010." No audio other than music. Shows a graph with lines appearing, on a monthly basis, showing how many were killed, and stolen years, building with each month until it reaches 9,595 killed, 413,342 stolen years. Music stops. Then, the music begins again and builds as it shows a graph and lines for 2013 for January and February, ending with 1,376 killed, 57,634 stolen years.]

So, in case you couldn't see it very clearly, what was happening was we were showing every single person who was killed in the U.S. in 2010 by a gun. And it starts off with a line. So for each person, they each get a line. It starts off as an orange line. And then when they die, the lines continues as a gray line, and we project what their life could have been like. And so, you see all of these lines build up over time.

So, I want to go into some of the decisions that we made going through this piece, just to show you that there could be hundreds of decisions that you make about data and how to represent it, but it doesn't have to be scary. They are all small steps that go into it. And over time, as you build your confidence using it, these should become not such scary topics.

So, first of all, out of the gate, that data was incomplete. So, that was the first thing we had to deal with was, okay, Florida doesn't report their homicides to the federal level. They just don't do it. Nothing against Florida, Alberta. They are wrong to not report it, in my opinion, but apparently they don't use the same methodology to report homicides, so they don't report it. But we felt that, you know, the data was coming from the FBI, so it's really the gold standard of data sources. You can't get much better than that. And it really covered about 95% of the population, so we felt, you know, that's good enough. We're going to note it in our methods and sources, and we're going to say that's fine by us. So, that was really kind of an easy decision to make.

The next decision we had to make was about age prediction. And in fact, it wasn't really a tough decision for us. It was sort of at the crux of the creative concept that it was all about the stolen lives. It was about the un-lived potential of these people. So, we knew we could get data. We ended up using World Health Organization data to project, predict what a life could have been. So, just like right now, I could [predict], you know, based on probability, how long might I live? What might my life expectancy be? We did the same thing for this data.

14th Annual International Symposium on Online Journalism

But it was surprising to us that after we released the piece, that was sort of the thing that people seemed to nitpick, including some people in this very room, including some people who invited me to this panel, in fact. [laughs/laughter] But it was great. It was a great question to ask. And I think that it really shines a light on how detailed you can get with data. It was, you know, a lot of people said, "Well, you should really show confidence intervals and this and that. And you need to be more accurate about those numbers." And to me, it was so clear that if we had done any of those things, we would have had to layer in a level of education into the visualization, and it really would have detracted from the simplicity of the piece, from the elegance of the piece. I didn't want to bog it down with a lot of explanation. We wanted it just to be clear and just to be very simple. A very simple expression of what we felt was being lost in the world. And so, it felt like, yes, you can get down to a level of accuracy, but even if we had gone into demographics, if we had gone into a lot of detail, I don't think it would have gotten us that much further. We wouldn't have gained all that much out of it. So to us, it was a creative decision to go with, you know, let's go with something simple that's maybe less accurate. We can live with that.

Another decision that we had to make was whether or not to use the 2013 data. Our first version just used the 2010 data that was supplied by the FBI. And we chose that route because we really wanted the piece to be indisputable. We wanted it to be bulletproof—pun intended. We knew we might face some attack by the gun advocates, so we wanted it to just be as solid as possible. But we also knew that Slate.com was doing some interesting stuff with gun data. After Newtown, they had started crowdsourcing gun deaths in the U.S. And it was really intriguing to us, and so we started looking at their data after we put this out. And [we were] thinking about, well, what's Version 2? Can we add this data?

And a lot of questions came up when we started looking at that data. The first thing being the data is completely different from the FBI data. Even though it's all about gun deaths, it was still there were so many differences to it. The Slate data included suicides and accidents, so the numbers are much higher. The data fields themselves are completely different. The FBI data has things like type of gun that was used, relationship to the victim between the victim and the killer. Was it manslaughter? Was the person killed by a cop? Those types of things. And the Slate data was completely different. It had other wonderful things about it, like individual names and things like that. So, it was different. It was hard to.... You know, we don't want people to make direct comparisons to the 2010 data.

And another problem we had with the Slate data was that it was crowd-sourced, which I love. I love that concept, but in terms of a data source and being able to cite that, it was really troublesome to me, because there were literally thousands of data sources coming in by thousands of writers/journalists. And the data was being gathered by a dozen people who I didn't know, [who] I'd never met. And it terrified me. [chuckles] As a data person,

14th Annual International Symposium on Online Journalism

that was really hard for me to stomach. And there's no way we could even cite that. I mean, if you go down into the database, you can get the exact source of those things, but that was really problematic for me.

So, what we ended up doing, we debated it internally a lot and discussed all of these issues, but then the way we actually decided was we put the data into the tool and we let it play. And that really decided it for us, because we all sat in the conference room and just watched it play out. And you know, I don't want to use expletives, but that was really f-ing amazing to see. Everyone was just silent afterwards, because you just watched, you know, what you just saw here. You watched those people. You're essentially watching everyone in America who has been killed so far this year play out. And you can go in and, you know, select a button that says 'yesterday,' so I can see exactly what happened yesterday. I can push a button that says 'last week,' so I know exactly how many people were killed last week. I know who they are, what their names are. I found, you know, Antonio that way. I can see more about who he was.

And finally, I'm running out of time, and I just want to touch on this briefly. The decision of what our focus would be for this piece, it was very clear from the beginning that we would do this about victims. We could have done it about anything. It could have been about gun control. It could have been about the killers. It could have been about domestic violence. It could have been on any number of things, but we really chose victims, because it was an emotional story.

We started building this in January and at a time when discussion of Newtown was starting to drop off the radar. And that was really sad to us that, you know, just as the gun debate was heating up, we were talking less and less about this tragic incident. And we really wanted it to be an emotional story. We wanted to grab people by their collars. We didn't want people to forget about Newtown and what happened there and how, you know, how we can all — we all need to make a change. We all need to do something to move this country forward.

And secondly, the focus on victims also served as a common ground, because it is indisputable. It's indisputable that Antonio should be here today. He should be here. There's nobody in the world who would say he deserved that, so it's a common ground. It allows both sides, everyone from anywhere in the spectrum to come together and say, "Yes, we agree on that. If there's one thing we can agree on, we can agree on that." And it really starts to be a spot where we can open up a discussion and we don't have to be bogged down by our party lines or by lobbyists and whoever else has a voice in this. It really allows for a common ground where we can start a discussion. And I feel like that's the importance of where data visualization lives.

Thank you.

14th Annual International Symposium on Online Journalism

[Applause.]

Alberto Cairo: I had prepared a couple of questions to ask the panelists, but I think that it would be better if you ask questions ... to us or to them. So, go ahead. Who wants to ask question. Have a question over there.

Dave Grimm: My name is Dave Grimm. I'm a Deputy News Editor at Science Magazine. And like journalists at a lot of publications, this question is directed to Hannah. We were very inspired by Snow Fall. And like a lot of publications, we said, "Oh, wait, now we want to do that." Right? [chuckles] So, the big question for us was resources. You know, how can we...? Clearly, that's something you put a lot of time and money into. And as a smaller outlet, we wondered, how much is it going to cost us to do maybe something not on the level of Snow Fall, but something similar. And what's the return for us? And I was wondering if you would just talk a little bit about resources, economics, time. What was put into it? And what do you feel like you guys got out of it?

Hannah Fairfield: Well, because we hadn't done anything like that before, and anytime you do something for the first time, it generally takes a little bit longer. At the same time, we also covered an election, hurricanes, all kind of natural disasters, and a lot of breaking news at the same time, so it's not like a big group of us were sitting cloistered somewhere just working on this. So, the exchange of ideas started early. We definitely did a lot of iteration. We used each other to exchange ideas. And we also built some things that didn't really work. And I think that that's good, because what we ended up building overall was something that didn't have a whole lot of mistakes in it, which meant that other people can look at it and think about how they can use that as a starting point. And at the end of my presentation, I showed a couple of examples that are very recent in which it didn't take a really large team to build those. The team at The Post was definitely larger, but the Microsoft thing that I showed and the White House one, those are kind of inspired by it.

And there are also pieces of Snow Fall that people liked. Scott Klein was telling me that one of his favorite little pieces was a tiny little animated thing of the air bag. That you're reading along in the piece and there's a little figure that pops up kind of in the C column, and then as you read about what air bags do and why skiers wear them in avalanches, all of a sudden, the figure which you thought was static reaches over, pulls a little cord, and pop, the air bag comes up. So simple to do. Really, really simple to do. Built that in... Graham Roberts built that, and he did a prototype of it in less than a day. So, that's the kind of thing that if you have a story that would really work with something like that, and you've got somebody who can play with that a little bit and can envision what it could work like, you can build that in less than a day. That's not a big resource thing.

14th Annual International Symposium on Online Journalism

And for other things, I think that taking either pieces of Snow Fall or pieces of thinking about how we use the reporting that we had to build pieces of this sort of immersive story. You don't need a lot of resources to build something immersively. You just have to think about it, instead of having a story with some stuff on the side. You have to think about it in terms of a much more linear story, a much more immersive [story]. How can I use the tools that I've got even if they're not big tools? But I've got tools. All newsroom do. How can I use that to bring that into the package? Instead of thinking about the data visualizations and a lot of things on the side.

One of the things I loved about the Washington Post package was that they actually had what Alberto would call a true classic data visualization that you clicked on and explored and got to play with. It was embedded right in the middle of the story right when you wanted it. You didn't have to click a different page to go to it. It was right there in the package. It was elegantly done.

Alberto Cairo: We have another question over there.

Question: Hi. So, Scott's heard me kind of talk about this kind of ad nauseam, but I'm going to just keep bringing it up. One of the things that I was really bummed about given our timeline when we were launching the Globe site is, we didn't get to a point where we set up some really great base files and template files for the graphics department to be publishing data visualizations onto a responsibly designed site. And part of the reason we couldn't get there is because we were trying to figure out whether or not it's a problem if the visualization — the URL to the visualization on mobile is different when it's on a small view board versus a large one. And so, I pose to you guys, can you please talk about, if you've started talking about doing data visualizations on mobile or responsibly designed sites, some approaches that you've had and thought about? Even if you haven't done anything, can you talk a little bit how you think you might tackle this?

Scott Klein: That that I was showing, the schools, the opportunity gap is responsive, so it has a — I'm not sure if it has an iPad view, but it zooms down to Smartphone. The rule at ProPublica, I mean, you know, we sort of jokingly say that if a news app that we make isn't responsive, it may be broken. We say that. We don't go all the way and say that it's broken if it's not responsive, because there are some visualizations that just can't work on the small screen. It can't sort of be folded up to work on a small screen and would require like swapping out the entire thing and put in another mobile version of it. And we actually did that once with a project we did with the Washington Post, and they contributed the mobile version for us. So, kind of by expanding the team, we were able to do it. But yeah, we do responsive as often as we can, literally, as often as we possibly can. More or less all of them now at this point.

Alberto Cairo: Another one over there.

14th Annual International Symposium on Online Journalism

Amalie Barras: Okay. Hi. This question is more for Scott and Kim. My name is Amalie Barras, and I'm actually graduating in May from the School of Advertising here, and I'm sort of pursuing like the analytics path. So Scott, you talked about how to make data more actionable in a way, like with what you were saying about how you zoom in to kind of a more local level, and how you can use that and know that you can cut down on your energy consumption if you want to change. And Kim, you talked about how — your presentation sort of brought it to life, like that very emotional aspect of, you know, the gun violence, for example. So, I wanted to ask, how can I in my career and how can we all as strategic communicators harness the passion that your work creates with the action ability of Scott's and like put that together?

Kim Rees: Scott and I could start working together and form the dream team. [laughs] I think you made some good points that there are a lot of sides to visualization and to using data and using information in these new ways. There is certainly the analytical side and showing things very clearly and concisely and accurately, and then I personally believe that the emotional side is — I wouldn't go so far as to say it's mandatory, but it's pretty darn close. I think that anytime you want to engage someone, it's you need to get down to that root of, what is that story you're trying to tell? What are you talking about?

I mean, I think that Scott's piece about the schools and showing that the poverty is pretty much, you know, depriving these children of a great education, that is such a tragic story. You know, if my company was doing a piece on that, we would get down to that level of, you know, let's look at what these kids have to go through. They've already been served the injustice of their socioeconomic status and now we're depriving them of an education in this country that's supposed to be so grand and the champion of all of these things. So, we would go down to, you know, from that data, what is behind the data?

And that's really where we start every project meeting. We start looking at the data. Then we start going back to behind the data. Where did it come from? What is it talking about? Who's behind it? Who are the...? I like to talk about data sources a lot. And it's not the person who's gathering the data and compiling it. It's the actual data source. You know, I'm 5'5". That data is mine. I generated that data. I'm the data source. So, you know, that's the way we like to look at data. That's always where we go back to is, where did that data come from? So, that's where I challenge people who are dealing with data. Please, don't just look at the file on your desktop. Don't just look at the numbers. Go back and find what's underneath it all.

Alberto Cairo: Do we have time for another question? If so....[points to Daniela Gerson for the next question].

14th Annual International Symposium on Online Journalism

Daniela Gerson: Hi. My name is Daniela Gerson. I'm with USC Annenberg. And I have a question about the ProPublica — it was called Opportunity Gap?

Scott Klein: Yeah.

Daniela Gerson: Opportunity Gap database. And also for all of you. When you compile these large databases that are with using annual data, do you build them so that you can update them? Like, I was looking at it, and it was from 2009, 2010, I think, and it was great data for a project that I could really use, but I would want some of the updated data. And I see it in a lot of these large projects. I'm curious if you build them to update them.

Scott Klein: We do. And in fact, it is a kind of ongoing kind of problem and opportunity for us, because so many of the news applications we do, like Opportunity Gap, I think we built in 2010, and then we updated it again with additional data. They didn't come out with new data, but they came out with additional data. We have done projects... We've been doing an ongoing project about the stimulus program. Do you remember that? Which was five years ago? Five years ago. And we've been updating it annually since. And when it's government data, you sort of have one set of problems. You kind of clean the data. Maybe the data changes in the nature, and you have to fix that. But we also do a project called Dollars for Docs, which tracks pharmaceutical payments to doctors. And that's a dataset that we put it together ourselves by scraping pharmaceutical company websites. And that can take — I think the last update took us eight months. So, yes, we want [to do that]. And we can think that it's part of our responsibility to keep the data up to date so that people can come back and see how things are tracking, but which projects we do that to, how we retire a project if it's gotten to be too old or too much work, those are ongoing things we're learning.

Alberto Cairo: Very good. Well, it's 2:59, Rosental, so we are going to wrap it up. And thank you so much. Thank you to all the panelists here. I'm going to ask you for a big applause for all of them.

[Applause.]