Harvard Data Science Review

A Conversation with Steve Ballmer

Steve Ballmer¹ Liberty Vittert^{2,3}

¹USAFacts,

²Harvard Data Science Initiative, Harvard University, Cambridge, Massachusetts, United States of America,

³John M. Olin Business School, Washington University in St. Louis, St. Louis, Missouri, United States of America

Published on: Jun 30, 2021

DOI: <https://doi.org/10.1162/99608f92.41803359>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

The *Harvard Data Science Review* got the chance to sit down with Steve Ballmer, the former CEO of Microsoft and current chairman of the Los Angeles Clippers. Steve was Microsoft's 30th employee in 1980, and retired in 2014 after 14 years as CEO. But what you may not know about him is that he and his wife Connie are working to help remove barriers to economic mobility for Americans living in poverty through their philanthropic Ballmer Group, and Steve started a nonprofit, nonpartisan organization called USAFacts, that curates and contextualizes government data online to help Americans ground their political debates in fact. With his well-known charismatic nature readily apparent, Steve joined us for a discussion on the future of data science in the America. The interview was conducted on July 23, 2019, in Steve's office in Seattle, by Liberty Vittert, Professor of Practice in Data Science at Washington University in St. Louis, a Royal Statistical Society (UK) Ambassador, and an Media Editor of HDSR.

This interview is part of HDSR's Conversations with Leaders series.

HDSR includes both an video recording and written transcript of the interview below. The transcript that appears below has been edited for purposes of grammar and clarity.

This interview originally appeared in the special print edition of HDSR's inaugural volume (2019), published 2020.

Visit the web version of this article to view interactive content.

HDSR | Steve Ballmer and Liberty Vittert

Liberty Vittert (LV): All right, just jumping right in, why did you start USAFacts?

Steve Ballmer (SB): After I retired from Microsoft and I was working with my wife on our philanthropy, we wanted to focus on kids in the United States who might not get much of a chance. Question is how much money is coming from government to help support the lives of those children? You say, well, the purpose of government isn't to bring in money and spend it, it's actually to achieve an outcome. Therefore, what numbers

can you use to describe the important outcomes of government? And that became kind of the anchor point, if you will, of USAFacts.

LV: These facts are just government data?

SB: I have my own way of thinking about numbers. I like to see things in their totality. You want things to add up to 100%—that way you can see the whole playing field. We wanted to use only numbers that came from the government. And we wanted to give people enough context. You can tell me A is double B, but if B is 1%, A is 2%, that gives a whole different context than just saying that something is double. So context, whole picture, government numbers, revenue expenses, and most importantly, outcomes, and that together became the genesis of USAFacts.

LV: You get in the data, create this context, and then you present it how?

SB: We create, literally, a 10K report for government in the United States. It looks just like a 10K that you would do for a business. Same categories. It's a complete description of the business. It tries to simplify government by breaking it into four logical pieces, all defined by the preamble of the Constitution: establish justice and ensure domestic tranquility, promote the general welfare, provide for the common defense, and secure the blessings of liberty to ourselves and our posterity.

LV: For the average citizen sitting at home, how would you want them to use USAFacts?

SB: It's really who we're designing USAFacts for. It's not designed for what I would call the cognoscenti, or the numbers wonks. We said look, we have to at least be as appealing and understandable as newspapers. We have a lot more data for people who really want to drill deep.

LV: For a specific scenario like the upcoming presidential debates, where do you see USAFacts' place in that?

SB: We are nonpartisan. We are not trying to say 'take approach A, take approach B,' or 'that candidate D is better than candidate E.' That's not what we do. But on each topic that will get framed in the election, we want people to see the numbers. These numbers shape part of the dialogue.

LV: So, in the vein of all information serves some interest or agenda, or goes against some other interest or agenda, it could be said that there's no truly objective data. How does USAFacts deal with that?

SB: Well, I think there is truly objective data. I think numbers are numbers. They are nonpartisan. They describe what happened. *Forecasts* are very partisan. You take any issue, and I'll find you an economist who will argue any side of the issue. But what happened? That's objective. You know, 36% of kids in eighth grade can read at grade-proficient level. What's to be debated about that?

LV: Is getting all that data from the government difficult?

SB: The government is a pretty open book about the data that's produced. We use over 100 government database sources. But, sometimes the data is conflicting. The data that's much harder to get is specific local and state data. They're more locked up. Even federal data, oftentimes we have to go rummaging around through PDFs. They're not all in nicely organized, you know...spreadsheets.

LV: Is there anything from government data that you would want, that you're not able to get?

SB: There are probably things we don't know about and will not be given. Right now, I don't think that's the number one problem. I'll give you an example. I think our citizens deserve to understand whether we are sustainable if a cataclysm broke out in the world. Are we energy and food sustainable? Could our country produce for itself?

LV: Are there ever unintended consequences, though, with collecting that much data?

SB: Well, we're not telling anybody what to collect. I mean if somebody wants our opinion, we can give them things that might be interesting, but policy and the will of the citizens should determine what gets collected. CIA spending is, for example, not reported. It's okay that government has deemed that improper to put out there. We went through this recently with the Census: should we collect data through the Census on who's a citizen and who's not?

LV: I want to go back to something you said earlier about how only 36% of eighth graders are reading at a proficient level. How do you have the public learn about these facts?

SB: We do specialized reports that take individual topics and try to talk about them on a level that is more approachable. What if I tell you that math scores are up from 15% in 1992 to 34% in 2017*? You could say hey, that's good improvement. You could say, hey, that number is too small. You could say the amount of money we're spending on education children per child has gone from \$8,600 to \$11,800.

LV: In the same vein, you figured out the way to get software in people's living rooms was through video games. So how do we get data into people's living rooms, not just for the youth, but also for adults who aren't necessarily going to go back to school?

SB: We continue to experiment. What is the role of videos, where can you show visualizations, but also talk about them? What is the role of simply providing the material in some online format where people can read? How much do people want to random access the data through search?

LV: What can we do in education to get young people excited about data?

SB: Just my opinion, the best way to do that is trying to put things in the context of other things people might be more interested in. Kids are interested in sports statistics. Can we use that interest to leverage interest in data? You get kids pretty jazzed up about doing robotics projects in school. Can we get kids interested in that same manner?

Another thing that is hugely difficult, is people don't really deal well in large numbers. The notion that overall government spending is something called \$5.9 trillion. Well, what is that? How do you get people to get their minds around those numbers?

LV: Not to put you on the spot, but do you have any favorite examples of how you explain big numbers?

SB: Sure, how big is Africa?

LV: I have no idea.

SB: Well, you can put, I think, India, Argentina, the continental U.S. and Western Europe, all inside of Africa. And I said that the first time somebody at Microsoft was trying to convince me to invest in Africa.

The Constitution says we must have a State of the Union address. I might say we should have a State of the Union by the numbers. I can't remember the exact quote, but I think it was Madison who said that if we try to have a democracy without informed citizens, it's kind of a farce.¹ Ultimately, I'd love to see government have to nominally portray itself. Politicians signing on the bottom line that says, 'I've read these, and they, to the best of my knowledge, represent what has happened.' Corporations have to do that, it might be a good thing for government.

LV: What is your intent for USAFacts, what do you want people to really know?

SB: What I would say to people is, 'Hey, look, if you really want to participate in the discussion and decisions about what should happen in our country, you've got to get yourself a little informed.' We're going to make that as simple for you as anybody else on the planet.

LV: Where do you sort of see the evolution of data in our country going? How will it inform the future of our country?

SB: Well, there are two different ways. Along the vector we're talking about, in general, government has to become more outcomes-oriented. You know, you spend money on something, you want to understand what the outcome is, and how to measure it.

If you think about data more generally, the ability to mine data in order to improve a variety of things that go on, it's clearly there. You'd certainly hear that out of the tech industry today, but you also see it in a variety of other industries. We want to make elevators run more reliably, there's a huge amount of sensor data that you

collect from elevators. I'm five years out of date on the elevator industry, but I knew something about it when I was at Microsoft. Okay, how do you use that data to make sure you're assigning repairmen, right time, right place? Is there data to predict where you're going to have faults, and issues, and problems? There's so much data, you have to figure out what is actually important and how you might actually use it.

I think there's a natural human desire to say, 'Let's collect everything.' And actually, that's relatively cheap nowadays, the computing resources needed to do that are not very expensive. So it might not be a terrible idea. On the other hand, I think you need to have a theory of how you might use the data. We talk about that in the context of USAFacts. There's a bunch more data sources that we can collect.

We want to sell what we built, and build what we want to sell. What does that mean? What do we want to really discuss with the population, and what topics are they interested in? But, we also want to collect data in other areas to see if we can find exciting and interesting patterns that might also be interesting to people.

LV: As you said, we can collect all of this data, but is there a point when the data and the facts don't tell you everything?

SB: Data and facts never tell you everything. I mean, just take abortions. We don't know the number of abortions that happen in the United States. Only some states collect the data. That's an incomplete data set, if you will. But [if we had all the data would that really] change your mind about whether you're pro-choice or pro-life? I don't think so.

So there certainly are decisions that are going to be made on your instinct, your values, your judgment. It's funny, I'm sitting here, 'Numbers, numbers, numbers, numbers, numbers, numbers, numbers, numbers.' But if you take sports analytics and numbers, when it comes right down to picking players, the numbers are not our primary thing. We look at the numbers, but there's a lot of judgment, you know? There's a lot of judgment.

Now, if you're trying to game plan, who should guard who, in what way, the statistics are very useful. But if you're actually trying to decide whether to draft, you know, John or Tom, data is not going to be all that useful. It's not going to be as useful as most people think. [It's] different in baseball.

LV: Has there been an experience where you had all the data, and it's telling you one thing, and your gut instinct, experience, whatever you want to call it, has made you realize that that's not the right decision?

SB: Hmm, not very often. Let me explain why. To me, data is about describing the world and telling a story, mostly. The numbers describe to me the playing field if you will, but they don't tell me what to do. I mean, I would say, in some senses, Tom Brady, when he's back to pass, is like that. He sees the playing field, he knows where x, y, and z are, he has a sense of how far they are. He has predictions on how this cornerback might defend that wide receiver. A lot of data, but at the end of the day, you're going to make a judgment. 'Can I

make that throw today?’ That’s kind of the way I think about my decisions too. I visualize through the numbers, but then, at the end of the day, you know, you’ve got to make a decision.

LV: Getting a little bit into this sports world and data, is there data science that goes into it?

SB: There are six cameras in the ceiling of every NBA arena. On top of computer vision, you can build a machine learning layer that actually says, ‘Ah, I recognize that. That’s the sixth kind of pick and roll you can run. That’s a blitz.’ Coaches will look through it, and they’ll say, ‘No, no, that’s not right.’ And then you have a continuous learning process about more and more plays, more and more moves, more and more everything else. That’s created an incredible analytics product for our coaches, and you can say, Blake Griffin and Reggie Jackson [both of the Detroit Pistons], this is how they ran pick and roll. What does it look like, what were the best ways to defend it? Did it depend on personnel?

If you want to, you can synthesize a real-time version of the game and show spectators exactly what’s happening. For example, you can show that the probability Kawhi [Leonard, of Steve’s team, the Los Angeles Clippers] makes this shot right now is 76%, etc.

LV: So you see the way people experience sports changing through the use of data science?

SB: Correct. I mean, I’m never quite sure when to say computer science or data science. Because at the end of the day, it’s a lot of programming, with a lot of data. The word data science has actually has become kind of a weird word.

LV: As a statistician, I don’t understand it either, so I’m right there with you.

SB: There are people who used to just be financial analysts, and they’re not called data scientists, and in a way, they are. There’s people who say, ‘I only deal with many, many petabytes of data’ or ‘I only do statistical analysis.’

There’s all these people, and they all legitimately can be called data scientists, but the words become a little ambiguous to me.

LV: Ok, some fun questions, if you had a magical genie that could grant you any dataset in the world, what would it be?

SB: This is not gonna be all that fun, if we could capture, and really reason and analyze over the data of kind of what happened in the lives of a bunch of kids who grew up in unfortunate circumstances, and what has really helped kids emerge out of poverty versus kids who didn’t? That’d be pretty interesting to me.

LV: Biggest regret?

SB: I missed my son’s appearance in the state basketball championship.

LV: Favorite ice-cream flavor?

SB: Favorite ice cream? I don't eat much ice cream, I will tell you.

LV: Come on.

SB: I'm always watching my figure, as they say.

LV: Okay, if you could eat as much ice cream as you want?

SB: Chocolate chip cookie dough.

LV: Most irrational fear?

SB: Fear of dogs.

LV: Really?

SB: Yeah. I'm getting over it, but I have an irrational fear of dogs.

LV: In what other career would you have liked to be a success?

SB: Well, I would have loved to be a professional basketball player (laughs). That one's not very hard, uh, I would have loved to do that.

LV: And are the Lakers going down this year?

SB: We have 29 other teams we compete with. We will be the best team that we could possibly be. We will compete with all we have, all our might. And we shall see what we shall see.

This interview is © 2021 by the author(s). The article is licensed under a Creative Commons Attribution (CC BY 4.0) International license (<https://creativecommons.org/licenses/by/4.0/legalcode>), except where otherwise indicated with respect to particular material included in the article. The article should be attributed to the author identified above.

Footnotes

1. [The actual James Madison quote](#): "[a] popular Government, without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or, perhaps, both. Knowledge will forever govern ignorance: And a people who mean to be their own Governors, must arm themselves with the power which knowledge gives." [↵](#)