Hello, Darla here. I am video chatting with two of my colleagues from the what former colleagues from The Washington Post Aaron Williams and Armando monument and they are members of the graphics team who just produce this really interesting dot density map showing racial segregation patterns across the u.s. So I want to talk to them a little bit about their process and how that. So my first question for you guys is how where did you get the idea to do this project? How did this come about? It came about maybe a year ago just over a hundred years ago. So somewhere in there. I initially was just reading a lot about gentrification and kind of implications for policy. And how many birds are made up and take that these are conversations Armada and I have also just had you know, like, you know, just on our own on a nigga to work and kind of the idea behind this part of it was you know, whenever you hear the term certification or even the word segregation it softened Road around but not a lot of you know to back it or a lot of the examples people talk about or anecdote. So I think one thing we wanted to try to do was be like, okay. Well we have access to some information. How about we try to map it and not just map of but also do some kind of analysis to see how that works or not divers in area is because they think about Transportation or any kind of you know, Anything around the idea of race and where people are a lot of it is objective. And so we wanted to see if we can bring a little bit of science no rigor to that exact topic. And so that was kind of excited. So where did you find the data that allows you to explore this question? So the nice thing about doing this project is that most of the data was already there. We relied heavily on the. Since I'm Census Data from the US Census Bureau and then in December of last year of the summer of 2017. The published The Five-Year American Community survey which included estimates and these aren't hard Collins like the decennial census, but they gave us our best guess of what the population is like or how the population has changed in the last five years. And so that allowed us to not just you. No historic census data, but then also kind of get to kind of like oh, well, what's happening now is happening today. The thing about census data is that it the geography changes a lot probably since his tracks and census block groups. So go you might overlap this free and since it's like just them as neighborhoods and people move around the city or region starts to develop and so we also got a lot of help from our. Had Melnick and and Keating who standardize all the geography of the data to the 2010 geography. So that's what allowed us to comparisons from 1990 up into the current is he has that up because how we just use the rocket itself. Who would it be there had double counts or something's going to been off. So he had already doesn't know really hard work of standardizing the data set at the geography level, which allows us to do our analysis much purple. That's great. So once you had an idea and you where to get the data, what was your process for actually creating the dots because I know this data doesn't come in dots right out of the box, right? Yes. It is not coming in dots unfortunately. It also doesn't come broken up by Race So in terms of the data on the side of the project we had to do six different categories of race, which is we base out. How most analysis that has been done on this one I could find and so, you know, we created six different groups and that required some pipelining data work on the back end just like making sure all of our data was right and then we took two approaches for generating the dots will actually take the same approach for generating the dots but two different ways of rendering them. I should say so in terms of actually how it was done. We took each block groups or block group just like of subsection of a census tract. So basically it looks like an area within neighborhood and look at the population counts for each race within that block and then we just ran a script that would place a DOT randomly in that block and then color it by whatever color we chose for race. The problem with doing that though is that if you just ran that scripts run without any kind of checks, it couldn't run forever because even though your red because you're randomly placing things about and a box and that's always get everything in that box. And so your script my run forever. So we actually said a hard limit which was I think like somewhere for like five or 10 times the population so say you have a population of which I did my math right operational hundred, you know, we'd run it thousand times. Whatever. Yeah. And then you know basically if we hit the number of dots in the geography hit the total population and the new or good Lon if we did it that many times but somehow the not get exactly the right amount it also move on and the part of the reason why we chose that is because in terms of when you zoom and look at the maps when I getting down to like the actual block level of the neighborhood is so pretty zoom out. So at that level you're going to have a lot of. Some poor girl ain't over and so in that case I was kind of a trade-off of choosing between what was accurate and what would allow us to publish right? Yeah, you know force will be at a more time and craft the computers. Maybe we try another approach but that was kind of emphasis in generating thoughts. And then we had our really nice beautiful static Maps as well as the interactive map. And so basically the way this process works is we generated a static image. Using node and some other libraries to generate these really big PNG files and we

didn't give them the illustrator to finish up and then we had another workflow that use a bunch of different tools to render tilesets using map the map box doctor how API and then took that day to put it on a map box servers and that's what's what powers the interactive. So it's the same pipeline terms of generating the dots and then how we actually render them at the end comes down to what we were trying to do. So for you know, getting those static perhaps was like you see in Chicago. We had one workflow that we recruit these really nice static maps in the mirror another word for it at the whole us and then uploaded it into that box. And and for the static maps for the static Maps because. We don't have to worry about interactivity and we don't really care as much how long it takes to process. You can have a much higher density of information. If I remember right the density of the static Maps was higher than in the interactive maps just because an interactive map can't display the as many data points even though Matt box Geo which is used to display. The interactive maps is super fast and using webgl technology. It just can't render dots on the you know capacity of like a hundred thousand per screen or something like that. Yeah. So we're there any along the way where there any particular like cartographic pitfalls that you ran into that kind of held you back in some way besides the density issue. Yeah. There were a couple that we came across first was because the census block groups that we were using. Cannot cut off at the coastline. They expanded slightly Beyond. So if our DOT algorithm is checking to see whether that we're in the census book group shape. It's possible that we could wind up with and we did wind up with several dots inside the ocean. No one lives there. Yeah, so we had lots of people like living in the ocean or just like off where the. Census block group doesn't draw its its shapes to match the coastline. So one of the work square and had this nicely drawn out node workflow came in and broke it off. And yeah, so to solve the issue of people living in the water we took we basically clipped the census block groups to a land Coastline shape file, which then gave us block groups that. You know did not have did not did not overlap with the ocean and there's still you know, if you look they're still rivers and lakes where occasionally, you'll get a DOT in there and there is no way we could get a dense enough shapefile. That would let us accommodate like. That sort of granular Geographic data, but you know given the time constraints of the project. Like Aaron said we eventually did want to publish it. So we figured that was a trade-off that we could live with it was very detailed the way it came out. Yes. Yeah another. If all we had was just being able to to display that many dots so, you know, even even displaying say what 10 dots per person like we're still talking about millions of dots and so we really had to optimize the tile layers and the data that was in the title are so when you're working with that many data points, like every little piece of data that's in there is important. So we really had to strip out everything possible and. We wound up separating each if I remember right Aaron we separated out each Year's worth of data into its own separate title set which led us overall like exponentially increased the density of the data that we could show on the map that is correctly as I think initially when we published the project we had one massive tile layer that included all four sets of data, so or. 1990 2000 2010 and then the 2016 estimates look really nice but we hires you levels will lose a lot of their identity. They're really kind of created this beautifully elegant fascinating wraps. So after we published the actually edit engineer mutual friend of ours from afar to reach out and say hey we actually can get more dots in there. And so basically a week after publishing our mind and I sat down with her. And we had a conversation and from there he and I went on not breaking up adjusting to work for the break out the tile sets of the separate the different years and different tile sets. And then when you're on the interactive map click a different year, that's when we fetch that new tileset so for and. And honestly when it was one big time site didn't act either ways and actually affect performance like this. It's so loaded about has to ask naturally Testament to how good that box is rendering data using webgl that he said because we were able to pack more bats at a per year level. Maybe it would give us a lot more Fidelity are so that's what's called the. Cool, but also being able to realize after answer after we had spent months. It's kind of like really trying to fine-tune workflow to have to work late and then another really interesting thing. I didn't came up was color, you know, particularly when you're dealing with these these colors really important and our mind and I really tried to like find. Data sent back to your counter all kinds of things like whether you're colorblind what happens when you talk all different layers at once and our mom is creating the color palette we have now, that's really nice. But that was after coming months of going back and forth. And of course my queen or categories. Yeah, it added a sixth and we added some other, you know them we get to just coming up with more and more common variables that affects a while to figure out yeah. That's really interesting. Well, thanks so much. I really appreciate you both talking with me. Thanks. Yes, of course. Good luck. Everyone. Yeah.