

A look at The Pudding's successes, experiments, and failures

Welcome back, everyone. This is our last video of the module. Video five. And in it, we're going to take a look at some of my most recent data stories at The Pudding. We'll talk through each project and show how each uses some of the techniques that you've learned about in the previous videos. We're also going to take a look at why it's important to experiment. And then we're going to look at some of our failed projects.

First up, this piece that you've seen probably several times or heard about several times in this video or this module is women's pocket sizing. So this is important because the project started with a single question. You know, why are women's pockets so bad? And that's because Amber and I were, you know, trying to fit our phones in our pockets almost dropped mine on the streets of New York, and we realized that so many other women are equally as frustrated with this phenomenon. We took the idea back to our male colleagues, and they were like, "Yeah, I don't know about this. I think it's one of those things that all women know about. It's a water is wet thing." They said the story was too obvious, but we kept pushing. And we said, "No, I think there's something there."

So in this piece, it employs one of our signature things at The Pudding, scrolly telling. It starts out with that and walks you through kind of our process. We hand measured 80 pairs of blue jeans. We went out to the stores and measured them, and we found the average size for both the men and the women's pockets and then compared them, showing that on average, pockets in women's jeans are 48 % shorter and 6.5 narrower than men's pockets.

This piece uses animation too in conjuncture with that storytelling to really walk the reader through step by step. After that, we walk through this interactive part, where you can click on different things that you would normally fit in your pocket, say your iPhone, and find out the percentage of pockets that we measured that could fit that item. That's 40 percent of women's pockets. If we look at a front wallet, 40 percent of women's pockets also fit that. A pen, that's 98 percent of women's pockets. A woman's hand, which you would think that your hand is supposed to fit in your pocket, but only 10 percent of women's hands can actually fit into women's pockets. Through those filters, we're also showing the different brands that we found that could fit your specific item. In this case, we're looking at a woman's hand, and this just gives the reader a little bit more information. If they're looking for specific brand, you can filter or sort. All 80 pairs of jeans. Yes. We measured each one.

Then we break into this section where we answer a few questions for people. We knew that people would have additional questions. And so we said, "OK, well, what about different styles?" Is there a difference between skinny and straight fit? And the average is not so much. And then what about back pockets? You know, women jeans are not as lacking in the back pocket department.

We'll wrap our women's pockets piece with a little bit of history. In that, you know, women's pockets have really been an issue for several centuries. And in fact, famed fashion designer Christian Dior in the 1950s said that men have pockets to keep things in, and women for decoration.

The entire design of this pockets piece is really evocative of denim itself. There's a lot of pieces that look like stitching with dashed lines. There's yellow and red tags that are really

symbolic of kind of the iconic Levi's tag on the back pocket of jeans. So the design really just pulls you into the subject matter.

We've also included some whimsical elements or Easter eggs, and in that are pocket in the piece at the top of it, if you click on it, it actually reveals this poem called Dangerous Coats by Sharon Owens that suggests that women's pockets could potentially be dangerous because we might be carrying seditious or improper materials in them.

The other thing that I should note about the pockets project is that if we scroll back down through it, you'll notice that there are no traditional charts in this piece. The entirety of the viz is pockets itself. And Amber and I intentionally did that because we wanted it to feel like pockets. We wanted the pockets to speak for themselves. And so we didn't feel like there were any additional bar graphs or line graphs that would make that argument stronger for us.

The next project that I want to take you through is the rise of hyphenated names in pro sports. And this is a project that I put out in May of this year. Again, like most of our projects at The Pudding it started with a central question. This time I was watching an NFL game featuring the player Clinton Dix, and his last name was emblazoned on the back of his jersey. And it kind of arched over his numbers because it was so long. And it really got me thinking. I feel like I've seen a lot more hyphenated names on jerseys recently.

So after that initial question, I went hunting for the data, and I pulled it from a ProSportsReference.com for leagues like Major League Baseball, and the NBA, and Major League Soccer. And then I created this heat map of players with hyphenated last names by decade. You can clearly see the trend that, yes, in fact, hyphenated last names are increasing. The most so in the NBA and then in the WNBA because women are often asked to hyphenate their names instead of taking a man's name in marriage. The WNBA stands out as kind of the outlier of all the leagues that I looked at.

This was one of those stories that started with that big to small narrative, and the fact that the first visual that you come to is kind of this overarching view. And then it slowly gets a little bit more specific.

The next visual you come to is this wall of names idea. And I had tried a lot of different ways to visualize this with more traditional charts, but like pockets, I felt like the subject matter made for the most striking data presentation itself. The first thing that caught my eye with this story was those hyphenated names on the back of jerseys. And I wanted to be able to see those names, and I wanted the audience to be able to see those names in the data.

So this creates like a slider or an animation to take you back through decade by decade of each of the names with hyphens, and then you can also filter league and see how that plays out in all of the different sports. As we scroll down, I can see that there's a ton of names or just a few. Only 14 hyphenated in the NFL in the 2000s. But the story gets even more specific.

So from the leagues we get into the reasons why people hyphenate their names. And through this, I look at players that are vignettes or symbolize some of the reasons. So for BenJarvus Green-Ellis, it would be that he combined his parents last name. And for Skylar Diggins Smith, of the WNBA, she married and combined names with her spouse.

This is another project like pockets that, you know, if we scroll back up through it, there are no traditional charts in it. And that's again, because I'm really letting the data and the story dictate the form instead of trying to fit a preconceived notion of what a chart should be into this piece.

The last piece we're going to take a look at is one called Colorism in High Fashion, and Amber and I at The Pudding worked at the story along with a freelancer like a Malaika Handa. And she looked at 19 years of Vogue covers to see how Vogue represents women of all shades. She used machine learning to identify female faces and then used a tool to filter out that background. Here we're using the scrolling telling method to walk you through this step by step. She averaged the color for those pixels that she identified and then removed any information about the color's hue or saturation, leaving us an idea with the light or darkness value for that color.

This piece in particular has two places where it's scrolly tells. And the second when you get to is with this bee-swarm chart, and it kind of takes you through what the data looks like from darkest shade to lightest shade and then transforms to another dimension by time. Then it takes you through the different periods of Vogue's skin-tone range of its models between 2000 and 2005. So definitely a tighter grouping. There's less distribution. That does increase as we get closer to present day. And then we switch to the faces here, and we did that because the faces of the models are a little bit more recognizable with their skin tones. But we wanted to use that animation to provide some object constancy there to set you up for the faces. To say: Here's the skin tone and then here's the faces in the exact same spot, so that the readers can make that connection.

One thing that we weren't really expecting when we looked at this data was a big reveal that even though Vogue's skin tones of cover models have gotten more diverse, that spectrum has fanned out. Really a lot of it comes down to one model, and that's Lupita Nyong'o. Here we look at the five darkest cover models, where three of those are Nyong'o herself, and then we compare that to the five lightest cover models, which have five distinct faces.

This project is one of those that's gets into what we do best at The Pudding, and that is to take a kind of cultural look at a topic that has a little bit more underneath the surface. We can call it like an iceberg where you're just getting the type of content, and then we're using this Vogue cover girl story to talk about issues like colorism, and racism, and tokenism. Or sometimes we call it like the spoonful of sugar approach where we're giving you, you know, giving the readers, meeting them where they are, a topic that seems non-controversial or are non-intimidating in the idea Vogue covers, and then taking them through a deeper topic through that process.

So now that you've seen three projects at The Pudding, you know that they all have their own distinct look and feel. A lot of them do use the technique of scrolly telling. Some of them don't have traditional charts at all, but most of them, you know, have a little bit of fun with them. And it's important to us to embrace the idea of experimentation. Again at The Pudding, we're not bound by the news cycle or harsh deadlines, so we do have the freedom and the privilege to do that and experimentation.

But we also kind of want to remind people who are possibly in those more stressful or time constrained jobs why experimentation is important. So first, if no one broke the rules, data

viz would be boring. It would all look the same. You wouldn't be able to tell one chart from the other. It keeps the readers and us interested. It's equally important that the people behind the story –that's us, that's you, data viz practitioners – are learning and engaged with the story just as much as it is the audience.

Experimentation also doesn't limit the storytelling by the delivery. It makes sure that the right form for the story takes shape. It doesn't constrain you, and it doesn't force you to try to fit a square peg into a round hole.

And finally, experiments add to the whole industry's toolkit. It means that we can do a lot of different things. It means that we challenge each other, that we increase our definition of what data journalism and data story telling can be, and it means that we try new things and engage with our readers more.

But how do you know when to break the rules? It's often a fine line that's kind of hard to distinguish. We're going to try. So let's talk about when to break the rules. First is, is there a better way to tell the story? You know, going back to that, don't fit the square peg in a round hole idea. Don't try to constrain yourself with the current definition of data storytelling. We want to always make sure that we're pushing that boundary to keep redefining what data storytelling can be. It's also a little bit more OK to break the rules if the story is more lighthearted. I touched on this a little bit in an earlier video, but experimentation is often much harder when you're dealing with serious topics. You want to treat those stories with the type of integrity and respect that they deserve, and sometimes trying to break the rules or trying to play it a little bit more fun is not the safest bet with those type of stories.

It's also OK to break the rules when you're working or looking for a connection with the reader. Ultimately, again, our job is to communicate information, and that job is made much easier when you have a strong connection with the reader. So if there's a way that you can invoke nostalgia or another really strong emotion in the reader, it's OK to kind of break and bend the rules then too.

On the other side, you know, you want to follow the rules when you're at risk of deceiving the reader. Often people take graphics and data visualization at face value. There's kind of this saying that charts don't lie, words do. But human perception is a powerful thing, and charts can be deceiving. So it's important to always remember the way that you're presenting information. You know, be very careful when breaking axes or creating non-zero baselines. Always use area when doing circle and bubble charts. You really want to adhere to some basic data viz principles so that you don't deceive or confuse the reader.

You should also follow some of those rules and principles when they're already heavily established and ingrained patterns. Here I'm thinking about the color blue being associated with Democrats and red being associated with Republicans in the United States. Historically, the colors weren't always this way, but right now they're synonymous with the two parties. If you switch them so that Republicans were blue and Democrats were red, you'd cause major confusion. And again, that's one thing you don't want to do with your stories is cause confusion.

Sometimes people have a similar argument, and when we're talking about the colors, blue for boys or pink for girls. And, you know, feminist data viz has really embraced the idea

that girls don't always have to be pink, and that's true. But we shouldn't do the reverse. We shouldn't all of a sudden make girls blue and boys pink. We should choose colors that aren't associated or aren't heavily ingrained in people's minds with those categories. So, for instance, girl could be teal and boy could be orange, and that would be safe instead.

So because we love experimentation at The Pudding, I want to kind of walk you through a couple of examples of our biggest experiments. The first one I'm going to start with is one called the Gyllenhaal Experiment. And this is one by Russell and Matt that took data from Colin Morris, one of our collaborators before, to kind of find how people spell stars' names. And this was an experiment in a couple of different ways.

One, the entire visualization is built on reader input, so we don't know what shape that data is going to take until readers answer the quiz.

And two, we're prompting users to spell names by listening to audio. So here I'm trying to spell Alanis Morissette's last name. I think I've got it. I'm wrong. And we do this for, you know, several different celebrities. And then we build these Sankey-esque diagrams in which you can kind of see the path at which people split off from the names.

For Zooey Deschanel, it's after the second letter. For Zach Galifianakis, people start getting bad at the third letter. One of the most contentious last names to spell was actually Matthew McConaughey. And he had over 800 spelling different variations because he had so many vowels in his name. You can't see all eight hundred variations here, but because they would take, you know, majority of the screen. But McConaughey, difficult man. Difficult name. Alright, alright, alright.

The next example of an experiment I want to take you through is this project on NBA defensive three seconds, and how officials and referees call this.

This was actually a project that was completely built out in code and then rendered in video. We had done this once before in a project where we partnered with the Smithsonian National Portrait Gallery, but that was for an exhibit on a big screen. And this one, we really wanted to pair it down and make sure that it was accessible to people off platform, off of our site, on social media. We were thinking specifically Instagram in mind. And we did this project this way because we realized that a lot of our pieces, a lot of the the most big robust data viz pieces, are internet specific. You go to a site. You have to be on that site to see them, and we wanted to try to reach a much broader audience with this piece. So we'll keep watching for just a little bit to give you a little bit more idea of what the video looks like.

The next experiment we'll take a look at is this one on what the path to fame looks like. Russell and Caitlin mined Wikipedia page views for artists and tracked just how they exploded over time. This is another large experiment website because the entirety of the screen is the chart in itself.

So as you scroll down, these different kind of snake charts come into play and trigger with the entire page as the axes. These are snake charts or worm charts or smoke puffs. We called them a lot of different things when we were developing them because they were a completely new chart type that Russell did using D3, where he encoded both color and the width of page views and the height of page view in this one snake-chart type.

The other thing that we did for this piece is you can see these little counter bubbles that count down on the side of it. There is no leading text on this piece. There are some subtitles in the bottom there, but the piece is completely narrated with audio. And as you scroll, you trigger different audio bits about the content that you're seeing.

This was one of those pieces where because it was so new that we really needed to make sure that people knew how to consume this piece. So we started out with a little bit more detailed chart on Cardi B's chart, which is the first breakout star was walking you through that. That gives you an example of what you're getting into. You can also view this in audio tour form, or you can just say I'm just browsing. And that kind of takes you out of that handholding step, so you're able to experience the article in two different ways. One on your own, and two with the guide.

And the last experiment that we want to look at is one again by Russell about the craft beer capital of the US. This one at the top says that we're in Austin, Texas, because it's geo locating us. So "hi" to all of you fellow Texans out there. But the thing that we really want to focus in on for the experiment is this section here.

And in this section, Russell created this rotating scatter plot here, so that users could input what was most important to them when determining what made the craft beer capital of the U.S. So if it was 100 percent at quality, see that Santa Rosa, California, comes out on top. If you wanted a 50/50 split of quality and quantity, you could see that San Diego is your best bet. And if the number of breweries nearby is your most important thing, Denver, Colorado, comes out as your top bet for the best city for craft beer.

At The Pudding, again, it's really important to experiment for us. We experiment a lot, and this is just a handful of experiments that made the cut. But tons of more experiments don't make it. And that's just a process of learning. They're not failures. They're opportunities to learn.

And then in our next section, we'll go into a little bit more about the opportunities for learning as I'll take you through some of our failed projects, or some of the projects that we never hit the publish button on.

So now we're going to talk a little bit about some of The Puddings's failures. And failures is a little bit of a difficult word to use because in a sense, they're not failures. Each of these projects has taught us something. The end result wasn't a tangible or sharable experience. And so we think it's important to share this, because for every published story that you see on The Pudding, there's at least one story that didn't make the cut. And that just means that the stories that you do see are more refined. We've spent more time learning lessons and building from our past mistakes to put out something that we're really proud of.

The first story I want to talk about is one that we called Speed Runs, and back in October of 2018, Matt and I, at The Pudding, started looking into world records for videogame speed runs. Did you know that you can beat Super Mario Brothers in under five minutes? Yeah, we didn't either, so we were intrigued. And we reached out to a guy named Summoning Salt, who owns a few records himself and makes super popular world record video progressions on YouTube. We'll take a look at one on Donkey Kong now.

After Rare made the famous Donkey Kong Country trilogy on the Super Nintendo, they knew they were going to have some big shoes to fill in the future. It's regarded as one of

the greatest trilogies of games ever made, and following it up on Nintendo's next console, the Nintendo 64 was not going to be easy. The result was this.

Donkey Kong 64. Reviews came in, and one thing that most critics commented on was the length of the game, with most saying that it took at least 30 hours to get through all of it. Once a few runners today look back at that 30 hour estimate, we laugh.

We got really excited because we thought that there was a data story here that in addition to these cool videos that Summoning Salt was doing, that we could show it in data. And we were really inspired by a piece we saw on The New York Times about Olympic performance and how that has plateaued. We thought we'd see something similar with speed runs, with the times dropping and dropping until there was no more room to improve. And so I used SpeedRun.com's API to start collecting the data, and off we went.

But those speed run plateaus looked really different, and we realized that was because with the Olympics, we've already narrowed down the field. You're looking at the best of the best, but anyone could submit a run to speed runs. So the spread of times for beating Super Mario Brothers ranged from that under five minute mark to anything like 20 hours. Still, we thought there was something interesting there. So we pivoted and looked at how the records fell over time. We were picturing this as some kind of waterfall chart or a drop type of chart. But the more we looked, we realized, okay, 538 did this, and they did it well. They did it about Super Mario, and they did kind of this drop fall chart.

So again, we were back to the start. We weren't really sure where this story could go from here. We weren't really sure the next angle we wanted to take was. So we've put a big pause on it and decided that maybe data visualization just wasn't the right medium for the type of story we wanted to tell, or the type of story that we were most interested in. We thought about it harder and that the Summoning Salt videos were engaging because they showed video footage of these actual tricks and hacks that people used to speed run through video games. And we realized that with our data, and with the data that was available, we really couldn't compete with that. So for now, we'll leave speed runs to Summoning Salt.

The next stalled project I want to talk about is one on political funds. I'm a self-described design nerd. I have a design type tattoo here and one behind my ear. And I was really excited about the idea that Republicans or conservatives would use more Serif based fonts than Democrats. After all, both the party and the font style is more traditionally conservative. And I really, really thought this would be the case just looking at a handful of logos. So in October 2018, again, because apparently this was the month of my stalled projects. I started to look, and I gathered the 80ish candidates running for the midterms that year, and the data by party for percent Serif fonts vs. percent Sans-serif fonts was a little bit underwhelming. It still showed that 31 percent of Democrats used Serif fonts, and Republicans used 20 percent of Serif fonts. But that was the flip side of what I hypothesized.

So then I looked at another angle. I was like, well, what if it's not necessarily a by party association? Because, well, I you have to look through independents, and the libertarians, and all these other factors. What if it's kind of an ideological slant? So I looked at how Serif fonts appeared in states that voted Democrat in the last presidential election versus states that voted Republican. And that split was even less evident with about 29 percent of

Democratic states, our Senate candidates, using Serif fonts, and then about 27 percent of Senate candidates in Republican states using Serif fonts.

My initial idea and one of the big catalysts for it was that there's been so much talk about how political design has evolved since the Obama election of 2008. In fact, there was an entire like 300-page book called *Designing Obama* that cataloged his creative use of the font Gotham and his O logo, which was kind of the window into America. And so then I thought, OK, well, maybe it's more of a temporal thing. I need to look at this over time to see if I could really find that kind of like before Obama after Obama split. But the data was just hard to come by. There's not a huge repository of a long time frame of all of these political logos. And so I put the project down again because I didn't have enough time to investigate exactly as I wanted to investigate.

Since then, there have been a lot of great pieces that look at adjacent things like the color that candidates use. And in fact, PoliticsAndDesign.com now does have a repository and a catalog of all of the political fonts and logos that we see in society today.

And the final project that I want to take a look at is something we called Florida Man. This project has the unique distinction of being dropped twice among our team. The first time, our idea was to look at Florida man headlines. Florida man, for a little bit of a backstory, is kind of this mythical, memeable hero. Almost like a man of Florida who basically goes out and does weird things, gets caught, and capers. Just has some very strange things happen to him.

And so we wanted to see if Florida man, or the idea of Florida man, was more extreme or more different than, say, Illinois man or California man. And Ilian and Caitlin began to research this idea and quickly found that Florida's open government laws contributed to the phenomenon of Florida man. Because Florida Sunshine Laws are very open, you can get just about anything you want open on the Internet. That includes arrest records. So we could easily get police reports of, you know, Florida man removes all clothes and jumps in a lake with an alligator. But we couldn't get that same detail for states like Illinois or California because their privacy laws and their reporting laws were much more locked down.

So after learning this, we stopped investigating the storyline, and the Florida Man story sat on our backlog for quite some time.

But just last month, Caitlin and I decided to approach this story from a different angle. This time we wanted to look at Florida man headlines and see if we could try to predict what he would do next, creating a calendar for Florida man's activities. Is Florida man more likely to do this activity on a Monday or a Tuesday? So we started doing more research and found, you know, a Florida man headline Twitter feed that we thought that we could pull headlines from, and we were really excited about the possibility of trying to predict Florida man's actions.

But really soon after we did the initial research, an article from the *Columbia Journalism Review* came out. And in this article, it basically stopped us in our tracks. It made us think about the actual people behind these Florida man headlines, many of whom struggle with addiction, mental health issues, and homelessness. And we were agreed that we might be falling into a kind of the tired, sensational media narrative, and that we weren't treating our data and the people behind it with the compassion and the respect that they deserved.

Shortly after the Columbia Journalism Review piece came out, the person who ran that Twitter account on Florida man ceased his operation because of the same reasoning. And I just want to take a step back and really think about our responsibility as storytellers in the broad sense, and that a lot of times when you're working with data, data can seem like numbers. It can seem faceless like you don't have a human connection to it. It can seem like there are no people behind these ones, and zeros, or these numbers. And that's the exact kind of wrong way to think about data storytelling. You should always remember how the data was collected, who the data speaks to, and who the data is about. Because that will make for one just a better story overall, and two a more responsible and engaged kind of post.

So, you know, again, there are lots of reasons that data stories don't work out the way you'd expect them to. The story can be better told in another format. The data doesn't match your back, your initial hypothesis. Or it's simply not right to tell the story. And again, that doesn't mean you failed. It means you've learned.

So that's a wrap. Thanks for hanging out with me this entire course. Hopefully over this module, you've learned a little bit more about data storytelling. What it looks like. What shape it takes. What makes it special. And what our responsibility is as data storytellers. Alright, don't forget, for your certificate, you need to make sure to participate in the discussion forums and take the quizzes. Also, be sure to check out the Facebook group for this course.

If you have any questions for me, post them in the questions for the instructors' forum. And if you have any questions related to the course platform, be sure to post those in the questions about the course platform forum.

For lots of tasty Pudding content, many of the stories that you've seen in this module itself go to www.pudding.cool. Yes, .com was taken, and we're just that cool. You can also find us on Twitter [@PuddingViz](https://twitter.com/PuddingViz), or on Instagram at [@the.pudding](https://www.instagram.com/the.pudding).

And currently we're doing a lot of cool experiments on Instagram and trying to repurpose our stories into the square format that you see there. So you can check out our stories on our website, and then see the reincarnation of them on our Instagram platform.

Again, thank you so much for watching. My name is Jan Diehm, journalist engineer at The Pudding.