GYMNASIUM

INFORMATION DESIGN AND VISUALIZATION FUNDAMENTALS

Lesson 5 Transcript

In-Depth Example: Mariano Rivera

ABOUT THIS HANDOUT

This handout includes the following:

- A list of the core concepts covered in this lesson.
- The assignment(s) for this lesson.
- A list of readings and resources for this lesson including books, articles and websites mentioned in the videos by the instructor, plus bonus readings and resources hand-picked by the instructor.
- A transcript of the lecture videos for this lesson

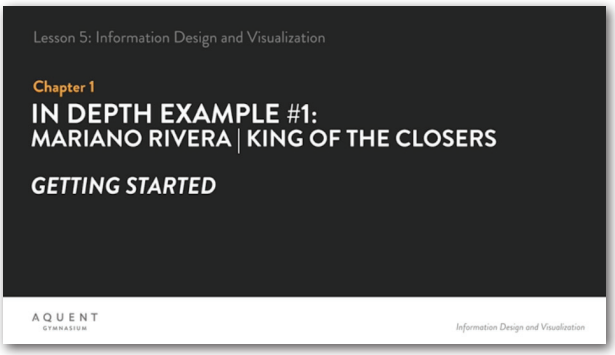
CORE CONCEPTS

1. The Mariano Rivera project is a valuable example to discuss because of the wide range of tools used to produce it.
2. The first step was to try to understand through the raw data available what type of information graphic needed to be made. It is important to interpret the data first before committing to a particular form.
3. Once the data was interpreted and information from subject matter experts was factored in, an initial prototype was developed in Adobe Illustrator. Prototypes are most valuable when they allow you to quickly identify what needs to be discarded and which elements might remain.
4. The initial prototype revealed that a form that emphasized a narrative would be best (specifically a video), so a script as well as a number of storyboards were created in order to begin organizing the flow of the piece.
5. Once the storyboard was complete, the assets needed for the imagery were created in a variety of 3D modeling and animation programs.
6. Once the assets were completed the data was imported into Maya, this ensured that the baseball models being used were accurate when animated.
7. When the data had successfully been applied to the animations in Maya, the project was moved to After Effects where multiple information layers were added as well as the final voiceover. Upon completion the project was rendered in After Effects into its final form: a video clip suitable for distribution on the Web.

ASSIGNMENTS

1. Quiz
2. Consider some alternate ways a graphic like this one could have been made using the data and resources at hand and create a plan of execution. When completed, post your plan on the forum and be sure to comment on one of your classmates forum posts as well.

INTRODUCTION



This is lesson five of Information Design and Visualization Fundamentals, an online course developed by Aquent. A few years ago, The New York Times published an information graphic project about the best closer pitcher of all time, Mariano Rivera. What's interesting about Rivera is that he has relied mostly on one pitch, the cutter, to dominate.

We've covered a lot of elements of information design, ways to think about it, best examples, software to use, et cetera. But, there's nothing like a real-world example to bring it all together. So, in this lesson, we'll take a look at how we developed "Mariano Rivera, King of the Closers", from conception to publication.

I think this is a good example to share with you because it uses a wide range of the tools we learned about in the previous lesson: data scraping to collect data; Excel, R, and processing to analyze data; Python to interpret data; Adobe Illustrator to prototype, and to storyboard; and Autodesk Maya, SmithMicro Poser, and Adobe After Effects to build elements, animate, and produce the visuals.

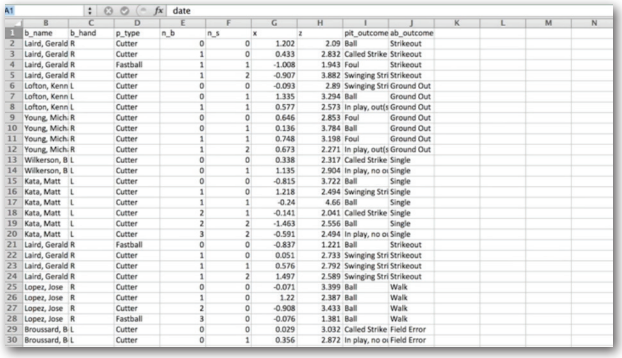
Before continuing, I recommend you take a few minutes and check out the project on the Times website. You can google, NYT Mariano Rivera interactive or follow the link shown, here. The link is available on the classroom site, as well.

As we learned in lesson two, choosing the story that's right for a visual presentation is the first step. In this case, the Times magazine was planning to publish a big piece on Mariano Rivera, and so, we were interested in the kind of visual reporting that might accompany this. We were aware that MLB.com had animations of pitches from each of the games, so we knew that there would be some data to play with.

In fact, PITCHf/x, a tracking system created by Sportvision, which was installed in every MLB stadium since 2006, collects data on every pitch thrown in the major leagues. For each pitch, they record the point of release, along with the position, spin, and velocity, with a series of images of the ball as it crosses the plate.

To collect and interpret this data, we teamed up with the sports statistics firm, Complete Game Consulting. They used data scraping methods to retrieve data for every pitch thrown by Mariano Rivera dating back to the 2007 season. This includes who the pitch was to, if they are a lefty or a righty, the types of pitches thrown, and the outcome, strike, ball, walk, ground out, et cetera.

Here's a sample of what the data looks like in Excel. A connected spreadsheet in Excel allowed us to enter a line from the spreadsheet, which would then reveal the data from the 42 images captured for position, velocity, and spin. We could then use this to plot the data in 3D space. Having a complete, robust data set like this was an exciting start, but our job as information designers was only just beginning.



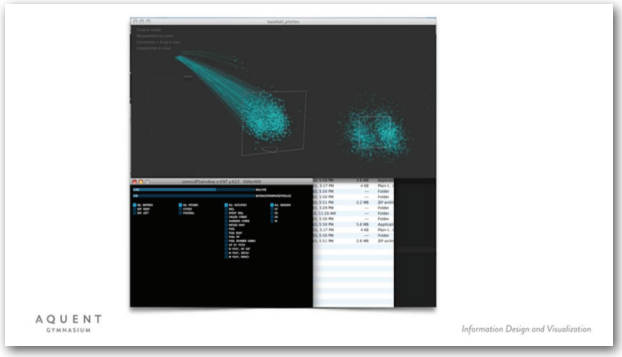
	b_name	b_hand	p_type	n_b	n_s	x	z	pit_outcome	ab_outcome
2	Laird, Gerald	R	Cutter		0	0	1.202	2.09	Ball	Strikeout
3	Laird, Gerald	R	Cutter		1	0	0.433	2.832	Called Strike	Strikeout
4	Laird, Gerald	R	Fastball		1	1	-1.008	1.943	Foul	Strikeout
5	Laird, Gerald	R	Cutter		1	2	-0.907	3.882	Swinging Stri	Strikeout
6	Lofton, Kenn	L	Cutter		0	0	-0.093	2.89	Swinging Stri	Ground Out
7	Lofton, Kenn	L	Cutter		0	1	1.335	3.294	Ball	Ground Out
8	Lofton, Kenn	L	Cutter		1	1	0.577	2.573	In play, out(s	Ground Out
9	Young, Mich.	R	Cutter		0	0	0.646	2.853	Foul	Ground Out
10	Young, Mich.	R	Cutter		0	1	0.136	3.784	Ball	Ground Out
11	Young, Mich.	R	Cutter		1	1	0.748	3.198	Foul	Ground Out
12	Young, Mich.	R	Cutter		1	2	0.673	2.271	In play, out(s	Ground Out
13	Wilkerson, B	L	Cutter		0	0	0.338	2.317	Called Strike	Single
14	Wilkerson, B	L	Cutter		0	1	1.135	2.904	In play, no ou	Single
15	Kata, Matt	L	Cutter		0	0	-0.815	3.722	Ball	Single
16	Kata, Matt	L	Cutter		1	0	1.218	2.494	Swinging Stri	Single
17	Kata, Matt	L	Cutter		1	1	-0.24	4.66	Ball	Single
18	Kata, Matt	L	Cutter		2	1	-0.141	2.041	Called Strike	Single
19	Kata, Matt	L	Cutter		2	2	-1.463	2.556	Ball	Single
20	Kata, Matt	L	Cutter		3	2	-0.591	2.494	In play, no ou	Single
21	Laird, Gerald	R	Fastball		0	0	-0.837	1.221	Ball	Strikeout
22	Laird, Gerald	R	Cutter		1	0	0.051	2.733	Swinging Stri	Strikeout
23	Laird, Gerald	R	Cutter		1	1	0.576	2.792	Swinging Stri	Strikeout
24	Laird, Gerald	R	Cutter		1	2	1.497	2.589	Swinging Stri	Strikeout
25	Lopez, Jose	R	Cutter		0	0	-0.071	3.399	Ball	Walk
26	Lopez, Jose	R	Cutter		1	0	1.22	2.387	Ball	Walk
27	Lopez, Jose	R	Cutter		2	0	-0.908	3.433	Ball	Walk
28	Lopez, Jose	R	Fastball		3	0	-0.076	1.381	Ball	Walk
29	Broussard, B	L	Cutter		0	0	0.029	3.032	Called Strike	Field Error
30	Broussard, B	L	Cutter		0	1	0.356	2.872	In play, no ou	Field Error

So, to recap, the Mariano Rivera project will be a valuable example to discuss because of the wide range of tools used to produce it. To begin, we started to collect data on his pitching with the help of Complete Game Consulting. We used data scraping to give us a complete set of data on every pitch Mariano threw over several seasons. In the next chapter, we'll take a look at how we worked with this data, to decide what kind of project we would pursue.

INTERPRETING THE DATA

In the previous chapter, we learned how this project came about, and how (and from where) we obtained the initial data that we would need to create an information design project that would try and explain why Mariano Rivera is so dominant.

So, now we had plenty of data to work with. We decided to narrow our focus on the 2009 season, in its entirety. We knew objectively that Rivera had a spectacular record, but our job as visual journalists and information designers was just beginning. We needed to now look at this data to see what it would show us definitively about Rivera's pitching, so we began a process of sketching with the data to explore it.

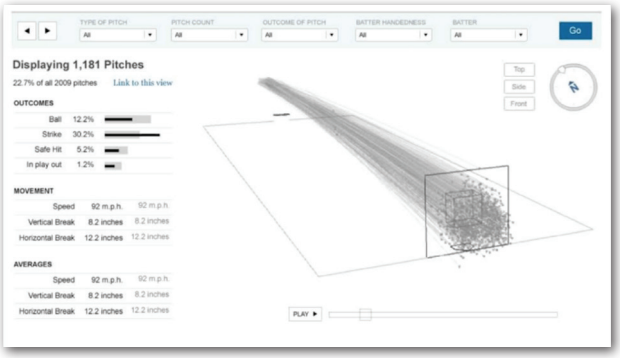


The data, being a path through space over time, was four dimensional. Each data point was a position in 3D space plus time. So we essentially built a data visualization graphic, to better understand what kind of graphic we should ultimately make.

We loaded the data into processing, which I mentioned in the tools chapter. This allowed us to explore the data in three dimensions, and to filter it based on a number of parameters, cutters to lefties, cutters to righties, fast balls to righties, et cetera.

We further sketched with the data in another tool we learned about, R. In R we created heat maps of where the pitches cross through the strike zone, to better understand the patterns in Rivera's pitching. From these, we learned some very interesting things, like his consistency in hitting the outside corners of the plate, and keeping balls out of the middle of the strike zone. We knew this was notable from comparisons with data analysis of other pitchers, who had much different strike zone patterns.

The data revealed a lot, but as journalists we have the opportunity to reach out to knowledgeable sources to get a more complete understanding of what we are looking at. So we visited John Flaherty at the YES Network. Flaherty had caught Rivera for years, and had a unique perspective on what made him such a great pitcher. He spoke of elements that the data on its own could never reveal. For instance, he emphasized the effortless nature of Rivera's pitching motion, which would often catch hitters off guard when the powerful pitch emerged.



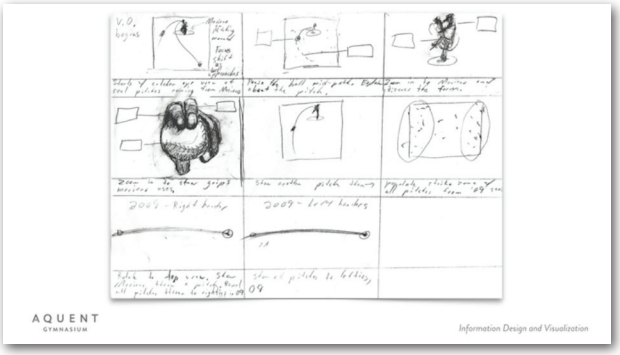
Having explored the data and conducted interviews with experts, the next step was to begin designing the graphic by prototyping. Using the visualization of the data from processing, which showed all the paths of the pitches at once, we mocked up an interactive graphic in Adobe Illustrator. This graphic would let users interact with the pitch data in a way similar to the processing app we built.

This is what that looked like; Readers would be able to swivel around the pitch paths, filter by righties or lefties, and by pitch type. There's something undeniably beautiful about seeing the data set in full, and playing with it.

But after spending some time with this design, we felt that something was wrong. It just wasn't the right approach. After all, if you aren't a major league pitching coach, it's very difficult to make heads or tails of the data, to understand what this data is telling us about why Mariano Rivera's pitching is so effective.

So, in the end, we realized that we should show readers what was interesting, rather than requiring them to find it themselves. Allowing a user to browse all of the pitches wasn't the most effective way of doing that. It's important to note that in information design, there will often be several iterations, and different approaches, before the most effective way of communicating can be found.

So, we changed directions entirely. After deciding on the main points we would want to highlight with the graphic, we decided on taking a more narrative approach to the graphic, which would involve writing a script, and beginning to design storyboards. We would try to give the reader an experience of the pitches, as if from the catcher's perspective.



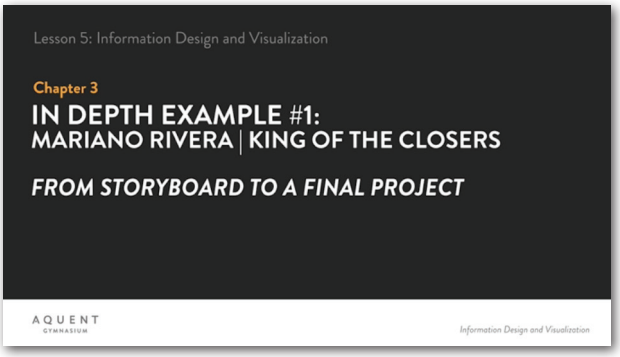
This is one of the early storyboards we created to figure out the flow of the piece. Pencil and paper, as lo-fi as it gets. And, the final product was not all that different from this plan.

So, to recap, once we had the data in hand, we narrowed the focus to just the 2009 season, began to sketch with the data, and conducted interviews to understand beyond the data. We then prototyped the project in Illustrator, but ultimately changed directions toward a narrative approach, which required the writing of a script, and storyboarding.

In the next chapter, we'll create the assets we need, and put everything together to make the final piece.

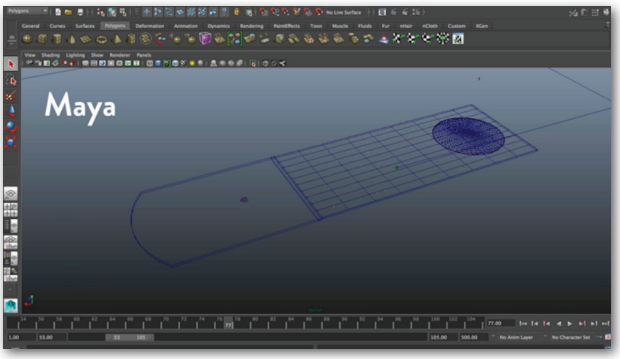
FROM STORYBOARD TO FINAL PROJECT

In the previous chapter, we decided on a narrative direction for the project, and narrowed our focus. We also created a rough storyboard to start organizing the flow of the piece. Now, let's look at how we created assets, and brought them all together.



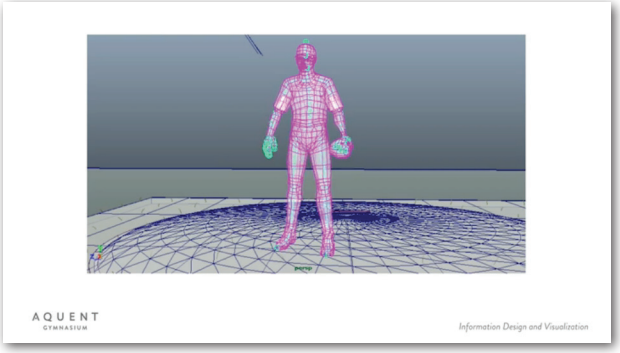
Having decided on a plan with our rough storyboard, the next task was to build the variety of assets we would need to create the imagery for the piece. We would need to create an accurate model of a baseball field that we could superimpose the pitch data onto. We would need a representation of the strike zone, and we would need representations of players.

To create these assets, we used the 3D modeling and animation software Autodesk Maya, as well as the figure modeling, and animation software, Poser. For the stadium, we decided to keep things simple, and only model the extent of what we would need to show the pitches. So, this was just the strip from the pitching mound to home plate.



For this, it was important to be accurate about the distance from the pitching mound to home plate, as well as the height of the mound. The most important element we would need was some representation of Mariano Rivera, himself, throwing pitches. Initially, we tried to animate the pitch motion, by hand.

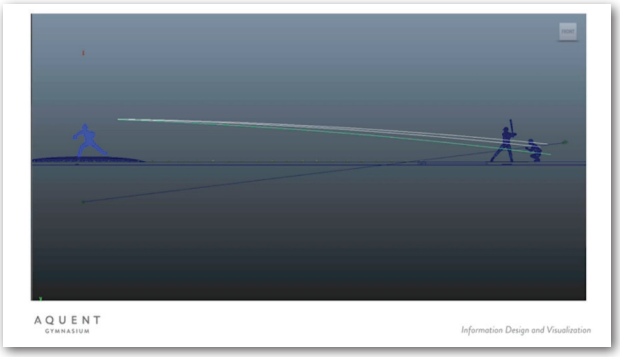
But, upon listening again to our interview with Flaherty we realized how important getting this pitching motion looking accurately like Rivera would be, to communicating what it would be like to face a pitch from him. We wanted to be confident that we were recreating his effortless pitching motion. We had recently been in communication with a motion capture studio at NYU called, The Movement Lab and so, we thought, hey, maybe we can get Rivera into the studio where we can put tracking markers on him, and record him doing his windup.



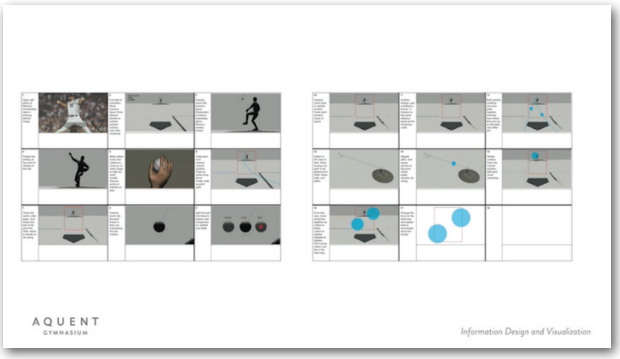
Well, it turns out Rivera is a busy man. So, we weren't able to get him in person. But, the movement lab had some technology to extract this kind of movement information from video. So, we sent them a series of videos of Rivera pitching.

They were able to track points in the video and use this information to reconstruct an animated pitch motion. We could then use this data to animate our 3D Mariano Rivera model. Now we had an animated pitch motion that we could feel confident was a close facsimile of the real thing, rather than our own made-up animated version.

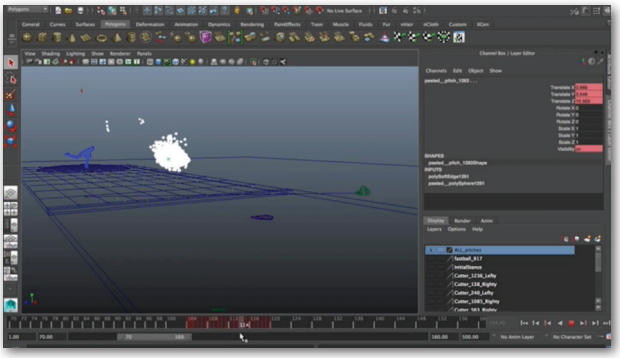
Having the main assets we would need to build the project assembled in Maya, we still needed to bring the data in. So far, we had only translated it in Processing. This is where Python came in handy. We wrote a script that would take the data points from each pitch, and draw curves in space in Maya. We could then attach a baseball model to each of these curves, and animate along the curve to recreate the pitch.



At this point, we had sophisticated elements all loaded into Maya, but we're still working with our initial, very rough storyboard. So using screen grabs of the elements we assembled in Maya, we built more fleshed-out versions of the storyboard that would drive the project using Adobe Illustrator. These contained more detailed descriptions of exactly what would be happening in each scene.

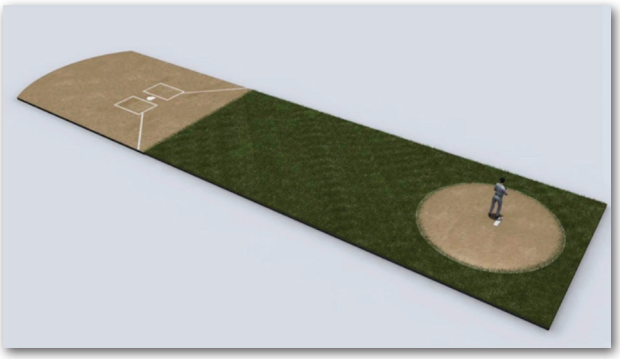


Now we were ready to animate. In Maya, we already had the animation for Rivera's pitching from the video-derived motion capture data. So, we were all set there. But, we needed to create animations of the particular pitches we decided to highlight.



In some cases, this was a single pitch. Sometimes, it was two pitches to explain how different pitches can look similar at first, but then, end up in very different places once they had crossed the plate. We were also animating all the pitches at once, and then pitches in different categories, such as all cutters to lefties, cutters to righties, and fastballs.

Next, there was animation for camera moves, for which we would use changing perspectives to show different elements, like focusing in on Rivera's grip. Lastly, there were animations of different kinds of baseball spins created for a section explaining what a fastball, cutter, and slider look like as they hurtle toward a batter. Once the animation was completed in Maya, we styled the scene, adding textures like dirt and grass, to give a more compelling aesthetic to the piece.



Rendering then calculates all of the lighting and textures in the scene, and outputs high-quality image frame sequences to be later assembled into video clips. These animated frame sequences were brought into After Effects, where we added the information layer. This includes an outline of the strike zone, labeling for the kind of pitch, and what batter it was, different spin types, as well as heat maps corresponding to the all-at-once pitch animations.

After Effects is also where the imagery was combined with a voice-over we recorded from the script. The project is then rendered out of After Effects, and we have our final piece.

So, to recap, we created assets in Maya and Poser, and created a pitch animation of Rivera focusing on accuracy. We brought the pitch data into Maya using Python, and updated the storyboard. We created the other animations we would need, and then stylized and rendered all the elements, which were then assembled in After Effects, with an added information layer. Then, everything was rendered out of After Effects to produce the final piece.

Now that you've shared in the journey from pitch, to data, through iterations, and to the final product, take another watch of the piece on the Times website. In the next lesson, we'll take a look at the thought process and tools used to develop an interactive information graphic from beginning to end. I'll be using a New York Times project we published called, "512 Paths to the White House", which uses data visualization and interactivity to present and explore Obama and Romney's chances of being elected, with analysis of the most competitive swing states.

Homework. The first assignment is a quiz. Each lesson in the course has one, and it's designed to help you reinforce the concepts covered in the lesson. Quizzes are available on the classroom site after this video is done. For assignment two, consider some alternate ways a graphic like this one could have been made using the data and resources at hand and create a plan of execution.

After you've finished your assignment, post a link to it in the classroom, and then find one of your classmate's assignments, and take some time to look it over, and offer feedback. That's it, for now. I'll see you in the next lesson.