I am an HCI researcher and data journalist
- Ph.D. from UW in 2021
- Work at NYTimes, FiveThirtyEight, CNN, New Yorker

Currently Head of Product & Design at Our World in Data, contributor to The Upshot at NYT
About Me

My CS research is inspired by working as a journalist
- Tools for interactive and data-driven storytelling
- Interactive research papers and educational materials
- Better decision making tools for policymakers
- Transparent and data-driven journalism on a deadline

Lecture will be biased toward data journalism but the material is more generally relevant (education, scientific publishing, policymaking).

@mathisonian
mathisonian.com
When Could the United States Reach Herd Immunity? It’s Complicated.

By Matthew Conlen and Charlie Smart  Feb. 20, 2021

With the vaccine rollout underway and coronavirus cases declining after a dark winter surge, it may seem as though the end of the pandemic is in sight. In reality, how soon could we get there?

One answer lies in herd immunity, the point when enough people are immune to the virus that it can no longer spread through the population. Getting there, however, depends not just on how quickly we can vaccinate but on other factors, too, like how many people have already been infected and how easily the virus spreads.

An estimate for the path to herd immunity

100% of population immune

80%
Motivation

Transparency in journalism
“A sense of the tentativeness of truth.”
- Philip Meyer, Precision Journalism, 2002

More engaged audience
Greussing & Boomgaarden, Digital Journalism 2019

Better learning outcomes
Mayer’s Multimedia Principles, 2005

Promote active reading
Victor’s “Explorable Explanations”, 2011

...also many challenges!
Topics

Storytelling
- How different mediums can be used to tell stories

Design of narrative visualizations & interactive articles
- Segel & Heer, 2010
- Hohman, et al 2020

Research on Authoring Tools
- Idyll, UIST 2018
- Idyll Studio, UIST 2021

Design exercise
Going forward
I carried wax along the line, and laid it thick on their ears. **They tied me up**, then, plumb amidships, **back to the mast**, lashed to the mast, and took themselves again to rowing. Soon, as we came smartly within hailing distance, the two **Sirens**, noting our fast ship, off their point, **made ready, and they sang**...
Narrative Storytelling

**narrative (n):** An account of a series of events, facts, etc., given in order and with the establishing of connections between them.

Effective storytelling “require[s] skills like those familiar to movie directors, beyond a technical expert’s knowledge of computer engineering and science.” - Gershon & Page ‘01
WASHINGTON — Justice Stephen G. Breyer, the senior member of the Supreme Court’s three-member liberal wing, will retire, two people familiar with the decision said, providing President Biden a chance to make good on his campaign pledge to name a Black woman to the court.

Mr. Biden is expected to formally announce the retirement at the White House on Thursday, according to one person familiar with the planning for the event.

Justice Breyer, 83, the oldest member of the court, was appointed in 1994 by President Bill Clinton. After the death of Justice Ruth Bader Ginsburg in 2020 and the appointment of Justice Amy Coney Barrett by President Donald J. Trump, he became the subject of an energetic campaign by liberals who wanted him to step down to ensure that Mr. Biden could name his successor while Democrats control the Senate.
WASHINGTON — Justice Stephen G. Breyer, the senior member of the Supreme Court’s three-member liberal wing, will retire, two people familiar with the decision said, providing President Biden a chance to make good on his campaign pledge to name a Black woman to the court.

Mr. Biden is expected to formally announce the retirement at the White House on Thursday, according to one person familiar with the planning for the event.

Justice Breyer, 83, the oldest member of the court, was appointed in 1994 by President Bill Clinton. After the death of Justice Ruth Bader Ginsburg in 2020 and the appointment of Justice Amy Coney Barrett by President Donald J. Trump, he became the subject of an energetic campaign by liberals who wanted him to step down to ensure that Mr. Biden could name his successor while Democrats control the Senate.
WASHINGTON — Justice Stephen G. Breyer, the senior member of the Supreme Court’s three-member liberal wing, will retire, two people familiar with the decision said, providing President Biden a chance to make good on his campaign pledge to name a Black woman to the court.

Mr. Biden is expected to formally announce the retirement at the White House on Thursday, according to one person familiar with the planning for the event.

Justice Breyer, 83, the oldest member of the court, was appointed in 1994 by President Bill Clinton. After the death of Justice Ruth Bader Ginsburg in 2020 and the appointment of Justice Amy Coney Barrett by President Donald J. Trump, he became the subject of an energetic campaign by liberals who wanted him to step down to ensure that Mr. Biden could name his successor while Democrats control the Senate.
STORYTELLING
across various media

PEOPLE TELL STORIES
WORDS TELL STORIES
IMAGES TELL STORIES
COMICS TELL STORIES
MOVIES TELL STORIES
DATA TELLS STORIES?
Lives of others

Jennifer Stoddard’s Canada Post job and her relationship with Facebook.

Jennifer Stoddard’s Canada Post job and her relationship with Facebook. She works at Facebook as a policy specialist, and in her personal life, she is also engaged in activism. She is concerned about the company’s handling of personal information and the potential for surveillance.

She has been involved in organizing protests and petitions calling for changes in the company’s data practices, including changes to the Facebook privacy settings. She has partnered with other activists to raise awareness about these issues and to push for more transparent and accountable data practices.

In her personal life, she is involved in various social justice movements and is active in her local community. She is passionate about protecting people’s privacy and ensuring that their personal information is handled with care.

She has been involved in organizing protests and petitions calling for changes in the company’s data practices, including changes to the Facebook privacy settings. She has partnered with other activists to raise awareness about these issues and to push for more transparent and accountable data practices.

In her personal life, she is involved in various social justice movements and is active in her local community. She is passionate about protecting people’s privacy and ensuring that their personal information is handled with care.
About 21,000 coronavirus cases were reported per day in early June, when the positive test rate was 4.8 percent.
**CASE STUDY** COVID-19 TESTING

**Input Data**
Daily case and test counts.

**Points of Focus**
1. Baseline (smoothed average of daily cases)
2. Counterfactual case curve
   (baseline positivity rate * daily test counts),
3. Actual case curve (daily cases)
Increase in Testing Does Not Explain Steep Rise in U.S. Virus Cases

As coronavirus testing has expanded, it is not surprising that more cases have been found. But they have risen much more than would be expected solely from additional testing.

Article and charts by state, Page A7.
Obesity Map Vadim Ogievetsky
1995
Obesity Map Vadim Ogievetsky
2008
Of the 7,666 times that police officers killed people in the U.S. between 2013 and 2019...
25 (0.3%) resulted in a conviction

74 (1.0%) resulted in a charge but no conviction

And...

Total disbelief about the system doing the bare minimum. George Floyd deserved so much more than this, he deserved to live.

Some detail about those 25 sentences:
Unknown sentence = 4 police killings
Just probation = 3
3 months in jail = 1
1 year in jail, 3 years suspended = 1
1 year in prison = 1
18 months in prison = 1
2.5 years in prison = 1
4 years in prison = 1
5 years in prison = 1
6 years in prison = 1
16 years in prison = 1
20 years in prison = 1
30 years in prison = 2

3 hours ago
Hans Rosling

“Best stats you’ve ever seen”
Narrative Devices

What devices communicates best?
Highly dependent on: audience, context, format

Format
Interactive Article
Animation
Lecture
Still Image
Video
CASE STUDIES

70% Journalism
20% Business
10% Research

Segel & Heer, 2010
Steroids or Not, the Pursuit Is On

Barry Bonds is going after the all-time home run record. He needs only six more to tie Hank Aaron and one to surpass him. Aaron

Lines are cumulative home runs

Hank Aaron
755 home runs
22 seasons

Babe Ruth
714 home runs
22 seasons

Barry Bonds
706 home runs
20 seasons

Bonds twice lead Home runs after 16 seasons
Ruth's 7th year
Aaron's 5th

According to allegations in a book about Bonds, he began taking steroids before the 1990 season, his 24th in the league. The two seasons after, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 36 in his first season of steroids. Here are projected home run totals for each player after age 34:

Aaron
Actual home runs over projected pace for five seasons

Ruth
Projected home runs over actual pace from age 36 to 38
Averaged 46.2 home runs from age 36 to 38

Bonds
From age 35 to 39, he averaged 46 home runs a season

Note: Ages are adjusted to July 1 of each season.

Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (11th), Rodriguez (7th), and Pujols (10th).

Hank Aaron
22 seasons

Babe Ruth
22 seasons

Barry Bonds
20 seasons

Willie Mays
19 seasons

Sammie Sosa
18 seasons

Pee Wee Reese
16 seasons

Mark McGwire
16 seasons

Ken Griffey Jr.
15 seasons

Alex Rodriguez
12 seasons

Albert Pujols
9 seasons

16 times HR 50 or mor p. H. L. M. M.
16 times HR 50 or mor p. H. L. M. M.
# Steroids or Not, the Pursuit Is On

Over the years, baseball has seen a trend of players attempting to hit home runs, with notable players such as Hank Aaron, Babe Ruth, and Barry Bonds setting records. The pursuit of hitting home runs has been a constant challenge, with players like Alex Rodriguez and Albert Pujols leading the way.

## Homer Pace After Age 34

If the accusations are correct, Bonds was 36 in his first season of steroids. Here are projected home run leaders for each player after age 34:

<table>
<thead>
<tr>
<th>Player</th>
<th>Actual home runs</th>
<th>Slugging percentage</th>
<th>Projected home runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonds</td>
<td>36</td>
<td>1.065</td>
<td>38</td>
</tr>
<tr>
<td>Ruth</td>
<td>714</td>
<td>1.045</td>
<td>67</td>
</tr>
<tr>
<td>Aaron</td>
<td>755</td>
<td>1.040</td>
<td>70</td>
</tr>
</tbody>
</table>

The data suggests that Bonds was in the lead, but the projections are based on the assumption that the accusations are correct.

## Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (517), Rodriguez (537), and Pujols (427), show how different paths can lead to the top. Each player has their own unique journey, with factors such as hitting, slugging, and defensive skills playing a role in their success.
Genres for Narrative Visualization (2010)
Genres + Interactivity + Messaging = DESIGN SPACE

**Author Driven**
- strong ordering
- heavy messaging
- limited interactivity

**Reader Driven**
- weak ordering
- light messaging
- free interactivity

STORYTELLING
- CLARITY
- SPEED

ASK QUESTIONS
- EXPLORE
- FIND

martini glass

interactive slideshow

drill-down story

Genres + Interactivity + Messaging = DESIGN SPACE
Communicating with Interactive Articles

60 Interactive Articles
NYTimes, WaPo, Distill, VisXAI, ...

Tied together research from
HCI, multimedia learning,
infovis, digital journalism

Hohman, Conlen, Heer, Chau, Distill, 2020
Five affordances of the format

Connecting People and Data.
Make data pleasant to work with. Happy readers are engaged readers.

Personalizing Reading.
Let readers choose the content that is relevant to their own experience.

Making Systems Playful.
Run interactive simulations directly in the browser. No setup required.

Prompting Self-Reflection.
Help readers learn by asking them to reflect in a low pressure environment.

Reducing Cognitive Load.
Use effective representations to make complex topics more intuitive.
Interative Articles - Design

Five affordances of the format

Connecting People and Data.
Make data pleasant to work with. Happy readers are engaged readers.

Personalizing Reading.
Let readers choose the content that is relevant to their own experience.

Making Systems Playful.
Run interactive simulations directly in the browser. No setup required.

Prompting Self-Reflection.
Help readers learn by asking them to reflect in a low pressure environment.

Reducing Cognitive Load.
Use effective representations to make complex topics more intuitive.

Most white boys raised in wealthy families will stay rich or upper middle class as adults, but black boys raised in similarly rich households will not.

Even when children grow up next to each other with parents who earn similar incomes, black boys fare worse than white boys, as seen in this New York Times interactive article.

Badger, Miller, Pearce, Quealy, NYTimes, 2018
Interactive Articles - Design

Five affordances of the format

- Connecting People and Data.
  Make data pleasant to work with. Happy readers are engaged readers.

- Personalizing Reading.
  Let readers choose the content that is relevant to their own experience.

- Making Systems Playful.
  Run interactive simulations directly in the browser. No setup required.

- Prompting Self-Reflection.
  Help readers learn by asking them to reflect in a low pressure environment.

- Reducing Cognitive Load.
  Use effective representations to make complex topics more intuitive.

When you were born, the Melbourne, Florida area could expect about 53 days per year to reach at least 90 degrees.

Popovich, Migliozzi, Taylor, Williams, Watkins
NYTimes, 2018
Interactive Articles - Design

Five affordances of the format

- **Connecting People and Data.** Make data pleasant to work with. Happy readers are engaged readers.
- **Personalizing Reading.** Let readers choose the content that is relevant to their own experience.
- **Making Systems Playful.** Run interactive simulations directly in the browser. No setup required.
- **Prompting Self-Reflection.** Help readers learn by asking them to reflect in a low pressure environment.
- **Reducing Cognitive Load.** Use effective representations to make complex topics more intuitive.

Sanderson and Eater, 2018
Interactive Articles - Design

Five affordances of the format

- Connecting People and Data. Make data pleasant to work with. Happy readers are engaged readers.

- Personalizing Reading. Let readers choose the content that is relevant to their own experience.

- Making Systems Playful. Run interactive simulations directly in the browser. No setup required.

- Prompting Self-Reflection. Help readers learn by asking them to reflect in a low pressure environment.

- Reducing Cognitive Load. Use effective representations to make complex topics more intuitive.

Buchanan, Park, Pearce, NYTimes, 2017
Interactive Articles - Design

Five affordances of the format

- **Connecting People and Data.**
  Make data pleasant to work with. Happy readers are engaged readers.

- **Personalizing Reading.**
  Let readers choose the content that is relevant to their own experience.

- **Making Systems Playful.**
  Run interactive simulations directly in the browser. No setup required.

- **Prompting Self-Reflection.**
  Help readers learn by asking them to reflect in a low pressure environment.

- **Reducing Cognitive Load.**
  Use effective representations to make complex topics more intuitive.

If job growth had been accelerating...

...the jobs report could look like this:

---

Irwin, Quealy, NYTimes, 2014
Challenges

**Few authoring tools available**
Those that do exist are too low level or lack expressivity
Not suitable for production usage

**Requires general purpose programming**
In practice need to be a web developer

**Difficult to evaluate**
Limited empirical evidence; hard to analyze usage data; lack of metrics

**Lack of design guidance**
Integration between text and media
Multiple output targets
How to make content accessible?
Idyll
A Markup Language for Authoring and Publishing Interactive Articles on the Web
DSL for Interactive Articles

Idyll Markup

[Header
  title:“Hi Idyll!” /]

This is the text of my article.

[Graphic
  value:x /]

And that was a custom graphic.

Input:
extended markdown
domain-specific graphics
DSL for Interactive Articles

Idyll Markup

[Header
title:“Hi Idyll!” ]

This is the text of my article.

[Graphic
value:x ]

And that was a custom graphic.

Input:
extended markdown
domain-specific graphics

Output:
javascript, html, css
DSL for Interactive Articles

**Input:**
extended markdown
domain-specific graphics

**Output:**
javascript, html, css

**ast**
- Component
  - type: Header
  - properties: title: "Hello Idyll"

- Text
  - content: "This is the text of my article."

- Component
  - type: Graphic
  - properties: value: x

- Text
  - content: "And that was a custom graphic."

**Live Web Page**
Hello Idyll
April 22, 2021

This is the text of my article.

And that was a custom graphic.
Text + Parameterized Graphics

Write articles in markdown.

Embed parameterized graphics.

Connect reactive variables to user input widgets to add interactivity.
Text + Parameterized Graphics

Write articles in markdown.

Embed parameterized graphics.

Connect reactive variables to user input widgets to add interactivity.
Text + Parameterized Graphics

Write articles in markdown.

Embed parameterized graphics.

Connect reactive variables to user input widgets to add interactivity.
Component Library

Input
- Check
- Click Me
- Radios
- Select
- Range
- Select
- Text Input

Presentation
- Chart
- Conditional
- Display
- Equation
- Git

Layout
- Aside
- Full Width
- Fixed
- Float
- Inline
- Scooter
- Stappare

Helpers
- 1,234 views
- <>
- Analytics
- Meta
- Preload
A JavaScript component API is exposed to end users.

Use with libraries like D3, Vega-Lite, Mapbox, Processing (P5), React, ...
Usage

Refined design through expert interviews.

Deployed with undergrad CS students at UW.

Students successfully in creating highly customized explorable explanations of algorithms.

Mathew, Cherukupalli, Pusich, Zhao, *Traveling Salesman Algorithms*, CS442 ‘17
Balancing Electric and Magnetic Force

Understand the basics for the velocity selector problem

By: Kitty Harris

Skills Involved
- Vectors: Addition
- Vectors: Multiplying by a scalar
- Particles: Sign of charge of electron + proton
- Magnetic field: Force
- Electric field: Force
- Forces: Balancing forces

Important Information

Objective
Tune the electric field so that electrons with a user-specified velocity feel no force.

Two rows and four columns, where each entry is a particle in coordinate plane where the electric field points in the x-direction and the magnetic field points in the y-direction. The particle is a proton in the top row and an electron in the bottom row. Velocities are as follows. First column: at rest. Second column: x-direction. Third column: y-direction. Fourth column: z-direction.

For all questions below, assume that the fields’ directions cannot be changed.

a. Consider the electron in the figure above. Four orientations of the magnetic field, magnetic fields E, and electron velocity v are shown. Draw or write down the magnetic force and electric force in each case. In which case could the electric and magnetic forces on the electron cancel?

You may experiment with the forces by changing the velocity in the 2D simulations below.

Click and drag the velocity arrow to change the direction of the velocity. The simulations represent the same system from different perspectives.
Idyll...

Seems great but...

...still requires learning and writing markup syntax

...requires use of general-purpose programming tools, e.g. to make a new post, run development server
Idyll Studio

Built a **structured editor** for editing and creating Idyll programs

Reduce use of general-purpose programming tools.

Eliminate a class of syntax errors.

Reify Idyll model.

Conlen, Vo, Tan, Heer, *UIST 2021*
Idyll Studio

Introducing UIST to the structured editor.

By: Matthew Conlen, Megan Vo, Alan Tan and Jeffrey Heer
Aug 11, 2021

Editing text

Any block of text that you see on this page can be edited. To edit it, move your mouse over a paragraph of text and click. You should see a black border appear on the left, indicating that the text is now being edited. Make the changes that you desire, then click anywhere else on the page to stop editing and apply your changes.

Text is written in a markup language based on something called Markdown. Markdown is a syntax that is meant to make it easy to add common styles to your text.
1. The interface enabled users to **rapidly create designs with less stress**.

2. But some users struggled to synthesize multiple concepts. Others **needed time to understand the reactive model**.

3. The interface **should provide more visual feedback**, promote experimentation, provide documentation on demand, and have **guardrails to prevent likely mistakes**.

4. The interface empowered users and **provided better support to less technical users compared to existing tools**.

<table>
<thead>
<tr>
<th>Role</th>
<th>Tech.</th>
<th>I1</th>
<th>D1</th>
<th>I2</th>
<th>D2</th>
<th>I3</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 comp. biologist</td>
<td>4.2</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P2 programmer</td>
<td>4.2</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P3 hci researcher</td>
<td>3.8</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P4 vis. practitioner</td>
<td>3.7</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P5 hci researcher</td>
<td>3.5</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P6 journalist</td>
<td>3.4</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P7 hci researcher</td>
<td>3.4</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P8 lab manager</td>
<td>3.4</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P9 programmer</td>
<td>3.2</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P10 vis. practitioner</td>
<td>2.9</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P11 editor</td>
<td>2.9</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P12 cs/journ. student</td>
<td>2.5</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P13 data scientist</td>
<td>2.5</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>P14 analyst</td>
<td>2.3</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>P15 designer</td>
<td>2.3</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>P16 writer</td>
<td>1.8</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>P17 ux researcher</td>
<td>1.8</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>P18 designer</td>
<td>1.3</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>n</td>
</tr>
</tbody>
</table>
Design Prompt
You want to keep the public as informed as possible of the status of the 2024 presidential election as it happens.

May want to convey:
- Current snapshot
- Likely outcomes
- Awareness of uncertainty
- Historical context
- Any unusual circumstances (e.g. delayed returns)
Design Prompt: ELECTION NIGHT

Input Data

Vote counts
- candidate, state, method, timestamp, # votes

Race calls
- candidate, state, timestamp

Constrains

- Near real time
- Never seen the data before
- Data may contain errors
- Should not erode public trust
Design Prompt: ELECTION NIGHT

Input Data
Vote counts
  candidate, state, method, timestamp, # votes

Race calls
  candidate, state, timestamp

Constraints
Near real time
Never seen the data before
Data may contain errors
Should not erode public trust

How would you visualize this?

What data would you use?
How would you visualize this?
What data would you use?

You want to keep the public as informed as possible of the status of the 2024 presidential election as it happens.

May want to convey:
- Current snapshot
- Likely outcomes
- Awareness of uncertainty
- Historical context
- Any unusual circumstances

Design Prompt: ELECTION NIGHT

Input Data
Vote counts
  candidate, state, method, timestamp, # votes

Race calls
  candidate, state, timestamp

Constrains
Near real time
Never seen the data before
Data may contain errors
Should not erode public trust