Open Source Your Data Design Process

Scott Murray
@alignedleft
alignedleft.com
Design is a process, not a product.
Design(ing) is a process, not a product.
Engineering is a process, not a product.
Design is a process, not a product, so you can't see it or point to it.
Projects Overview

Although my thesis process revolved primarily around one major project (Practice) and a minor one (Cheeky), these projects cannot be understood in a vacuum, as they were preceded by so many other projects. All of my prior projects at DMI have influenced my thesis projects, so I have elected to include descriptions of the most relevant ones here.

Data visualization has been an important theme in my work, and I discuss several projects, ranging from Search Explorer, an interactive tool for voyeuristic exploration of web search queries, to Relationship Visualizer, a tool for visualizing connections in network data. My iTunes Library Visualization was also a valuable learning experience, as was the BART Trains Visualization, which taught me that innovative designs are not necessarily successful designs. I briefly explore how data might sound—call it data auralization—with two projects, Audible Particles and Aural Data Plot.

With Dictionary Words and Questions & Answers, I explore elements of interactive, nonlinear narrative structures, to varying degrees of success. The Gesture Project and ASCII Photo Booth introduced me to the engagement potential of visual mirroring. Anticipation Study, my very first project created with Processing, continues to inform my thinking about how to communicate without words, and how to imply what is not yet seen.

Goals of This Document

In the pages that follow, I document and reflect on my personal process of engagement with dynamic media, concluding with in-depth analyses of my thesis projects. By the end, I hope to have shared the most pertinent and valuable insights and discoveries made during this journey at DMI.

My hope is that this document will be accessible, enjoyable, and valuable for anyone interested in design and interactivity.
Designer
person

Design(ing)
process

Designed
product

Outcome
result

(hopefully the intended one!)
Data design is a specific kind of design process.
DATA DESIGN

Any design process that is heavily informed by data, such that the output of the process (designed product) changes significantly when the data informing the process change.
DATA DESIGN

Any design process that is heavily informed by data, such that the output of the process (designed product) changes significantly when the data informing the process change.

DATA VISUALIZATIONS

Visual solutions to data-driven problems
Some Luminaries in Kindred Britain

Kindred Britain

Selfie City

Poem Viewer
Q: How do they do it?
Q: What data design process results in the most successful outcomes?
Data → Outcome
Q: What data design process results in the most successful outcomes?
Q: What data design process results in the most successful outcomes?

A:
Q: What data design process results in the most successful outcomes?

A: It depends.
Some factors on which “it depends”:
- team size
- team expertise
- project goals
- data availability
- output medium
- time available
- budget available
- tools available
Process Maps That Attempt to Clarify the Complexity of All This
- Interaction by typing successive numbers meant that the colors had to be modified in the visual refinement step to show a slow transition as points in the display are added or removed. This helps the user maintain context by preventing the updates on-screen from being too jarring.

![Diagram of data processing stages](image)

**Figure 1-12. Interactions between the seven stages**

The connections between the steps in the process illustrate the importance of the individual or team in addressing the project as a whole. This runs counter to the common fondness for assembly-line style projects, where programmers handle the technical portions, such as acquiring and parsing data, and visual designers are left to the end. The interconnection of these fields is more integrated.
Figure 10.59 Moritz Stefaner’s project workflow. (See the interview for an explanation.)
START HERE

What data do you have?

None or next to nothing

Best data ever

What do you want to know about your data?

Don't know yet

Find related data

New questions arise

Bar chart

Pie chart

Treemap

Line plot

What visualization methods should you use?

Scatterplot

Explore different dimension

What do you see and does it make sense?

(Nathan Yau. Visualize This. Wiley, 2011.)
Process diagram, Jeff Heer, University of Washington
A data journalism workflow, Mark McCormick with Simon Rogers, 2013.
A data journalism workflow, Mark McCormick with Simon Rogers, 2013.
Internet connection to the browser. My change to the structure of the data allows the points to appear slowly, as they are first read from the data file, employing the data itself as a "progress bar."

Interactivity by typing successive numbers meant that the colors had to be modified in the visual refinement step to show a slow transition as points in the display are added or removed. This helps the user maintain context by preventing the updates on-screen from being too jarring.

The connections between the steps in the process illustrate the importance of the individual or team in addressing the project as a whole. This runs counter to the common mindset for assembly-line style projects, where programmers handle the technical portion, such as acquiring and parsing data, and visual designers are left to choose colors and typefaces. At the intersection of these fields is a more interesting set of properties that demonstrate their strength in combination.

When acquiring data, consider how it can change, whether sporadically (such as once a month) or continuously. This expands the notion of graphic design that's traditionally focused on solving a specific problem for a specific data set and instead considers the meta problem of how to handle a certain kind of data that might be...

The answer to each question depends on the answers that come before, and it's common to jump back and forth between questions. As shown in figure 4.1, it's an iterative process. For example, if your dataset is only a handful of observations, this limits what you can find in your data and what visualization methods are useful, and you won't see much.

 acquisitions
 cleanings
 integrations
 visualizations
 modelings
 presentations
 Dissemination

 Figure 4.1 The iterative data analysis phases

 On the other hand, if you have a lot of data, what you see when you visualize data may be a mystery about other dimensions, which is why...

 do we know when the project is done? At this stage, we don't think about the data or the visuals, just about challenges, goals, and strategies to overcome the former and reach the latter.

 The second thing I need is data in raw form. I sketch many preliminary graphics with the data just to get a grip on it. I do these explorations with standard tools, such as Excel and Tableau, depending on the data set. I usually try to produce at most a dozen of sketchs.

 Figure 16.59 Moritz Stefan’s project workflow (See the interview for an explanation)
- Interaction by typing successive numbers meant that the colors had to be modified in the visual refinement step to show a slow transition as points in the display are added or removed. This helps the user maintain context by preventing the updates on-screen from being too jarring.

**Figure 1-12. Interactions between the seven stages**

The connections between the steps in the process illustrate the importance of the individual or team in addressing the project as a whole. This runs counter to the common fondness for assembly-line style projects, where programmers handle the technical portions, such as acquiring and parsing data, and visual designers are left to refine and interact with various parts of the data.
Decisions, decisions, decisions...
Clarify
What & Why

Explore
& Sketch

Define
& Produce
Acquire  Parse  Filter  Mine  Represent  Refine  Interact
Expertise is the informed capacity to make decisions.
Q: What expertise, exactly, do we need?
Q: What expertise, exactly, do we need?
Q: What expertise, exactly, do we need?

A: It depends.
EXPERTISE NEEDED FOR DATA DESIGN

ESSENTIAL

HELPFUL

BONUS
More people with more skills working more closely together making more of the right decisions more often.
PROCESS MAPS FOR BIG-PICTURE GUIDANCE

acquire → parse → filter → mine → represent → refine → interact

MORE EXPERTISE FOR BETTER DECISIONS

Mathematics → Non-defensiveness → Journalism → Visual Design → Data Fluency

BETTER COMMUNICATION FOR FASTER ITERATION
Open Source Your Data Design Process
Data Vis Process

If you practice data visualization — broadly defined — then I’m interested in your process. Share your experiences here, and I hope to synthesize the responses into some sort of cohesive taxonomy of process. I suspect our processes are quite different (by person, team, project, and context), but there may also be similarities. More importantly, we probably have a lot to learn from each other.

In the spirit of open-sourcing our processes, some of the data collected here will be shared back with the community. You may choose to answer anonymously, and I will ask for permission to share this information below. By completing this survey, you’re giving me permission to use this information in my own research, and to quote from it publicly (but anonymously) in workshops, presentations, lectures, and written work.

All information is optional, but appreciated!

Scott Murray
Assistant Professor, Design
University of San Francisco
http://xarts.usfca.edu/~shmurray/

Formal Title
What title is on your business card?

Layperson Title
What name do you use for your occupation when talking to people outside your field?

Dinner Party Scenario
At a dinner party, what do you say when the other guests ask what you do?
For you, what constitutes "success"?

"Client is happy (firstly); design is clean, code doesn't suck from my perspective, vis shows something interesting or allows people to find interesting things in it."

"Delivery on time. Actual usage by readers of the story. [Process produces] reusable code or components."

"There was nothing more to improve given the constraints."
First steps in the data visualization process:

"Collect raw data."

"Develop [a] hypothesis about the problem."

"Receive the data from client. Sit with client and make sure data is fully understood."

"I create folders for any idea that pops into my head. When I have time I dig into one and see if I can find a data set that fits my needs."

"Have questions to answer with data. All my projects start by having at least one question and some data."

"I usually start with a question. What do I want to know? What do I want to figure out? It usually starts with some curiosity…"

"ESTABLISH THE VISUALISATION’S PURPOSE"
1. What problem are we trying to solve?

2. What question(s) do we have?

3. What data do we have?

4. What (additional) data would we need to: solve the problem? answer the question?
Data Vis Process Research

github.com/alignedleft/data-vis-process

> take the survey
> read survey responses
> related reading list
> tons of links on process
Hugh Dubberly

“What can Steve Jobs and Jonathan Ive teach us about designing?”

1. Whole systems thinking
2. Deep, broad teams
3. Design conversations
“But only one form of conversation leads to a partnership, to deep trust, and ultimately to innovation and a sustained period of good design. Such conversations are principally about goals — about beliefs, about values, and about quality…”

– Hugh Dubberly
Open Source Your Data Design Process

Scott Murray
@alignedleft
alignedleft.com

github.com/alignedleft/data-vis-process

Today  2:30pm  Office hours
       3:15pm  Book signing