

THE ARCHITECT'S GENOME

MAPPING ARCHITECTURAL PAST, PRECEDENT, AND FUTURE

BENJAMIN BARBA | LEZHI LI | NICOLAS ROSSENBLUM | JAKE RUDIN



PROCESS BOOK

INTRODUCTION

Welcome to The Architect's Genome. This project is an exploration of data visualization in relation to the architecture profession. The architectural profession is categorized by styles, people, projects and their influence on one another, our goal is to begin to understand architects and the profession as a series of connections across time, using information about universities, locations, firms, exhibitions, and publications to visualize this complex network.

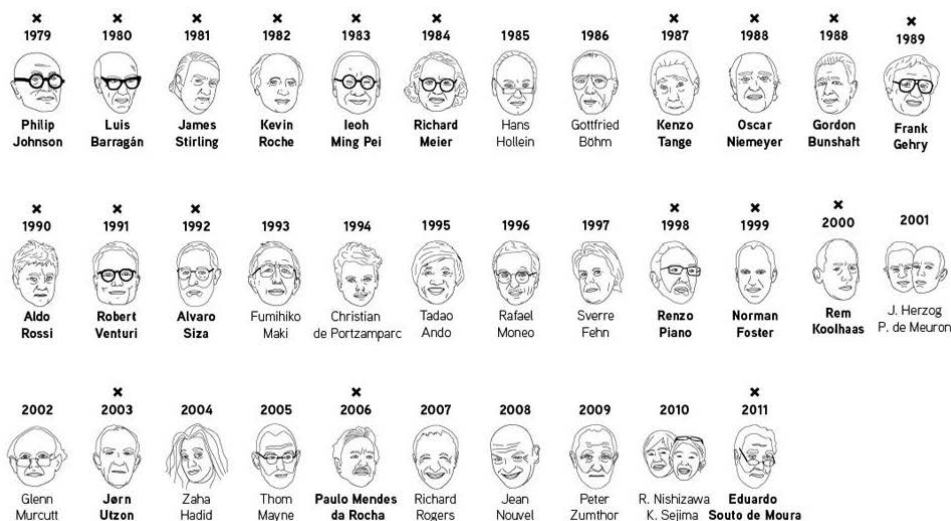
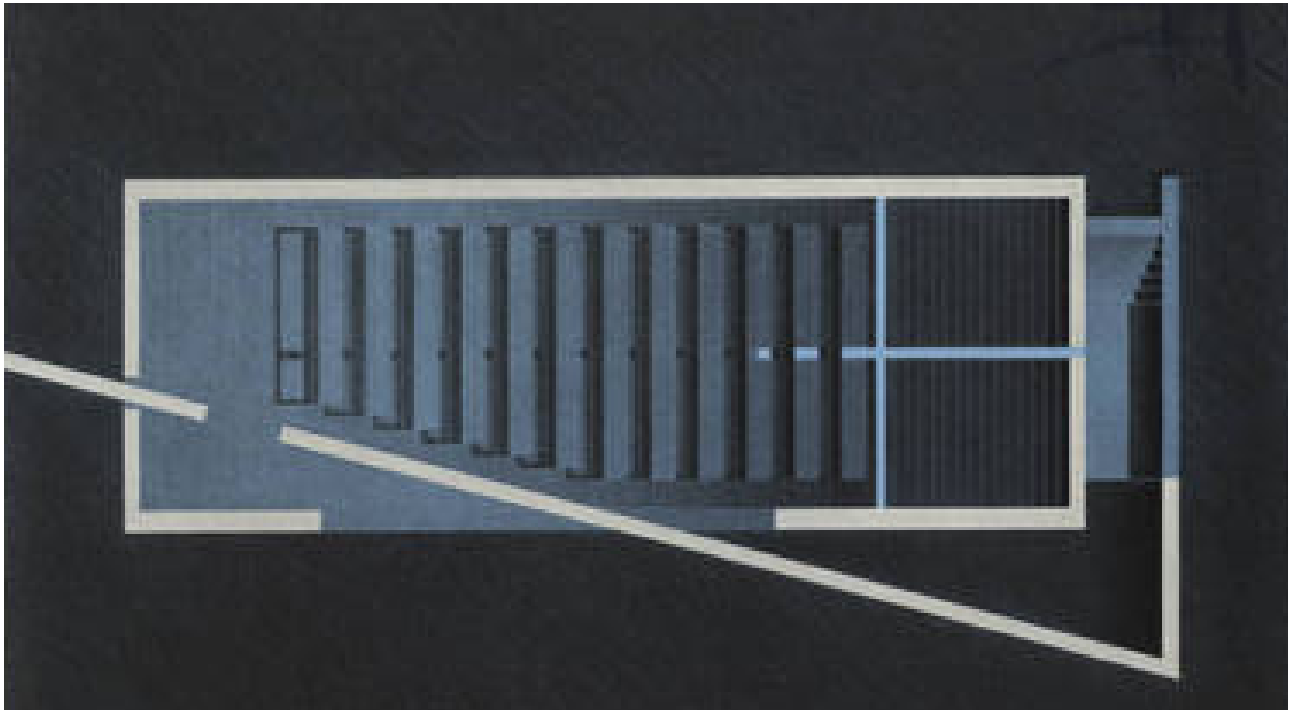


fig. 1 Pritzker Prize winners and the year of award

OVERVIEW AND MOTIVATION

With two members of the team in the Graduate School of Design and another member studying architecture within Harvard College, the team is motivated heavily towards understanding visualization as it pertains to larger questions of design studies. The education of an architect consists of a variety of skills, techniques, and information. Not only in school but in the professional world as well, architects find inspiration through researching precedents.

The problem is that the wealth of interconnected data regarding architectural history (relationships of architects to significant projects, exhibitions, publications, and other architects) has never been developed into a single, cohesive, beautiful and easily navigable database. We propose to visualize some of the information regarding architectural history and allow for interactive navigation of the relationships between architects across time. This project is in part a continuation of prior research done at the GSD and is a large component of ongoing thesis research by one of our members.



A BIT OF THEORY

Architecture, that which defines the built environment through the accurate and precise manipulation of light, space, and form, cannot truly exist without context. This is not to say that architecture cannot be, as it often is, irreverent of its site, neighbors and period, nevertheless it is always responding to the environment and the setting simply by existing within the two. Just as a construct is inherently affected by location (inclusive of time and surroundings), architects find themselves similarly inseparable from their backgrounds. The architect's genome contains the full complement of information he or she inherits from their experiences, whether the experiences relate to time spent at a university, working at an office, teaching, running a firm, or otherwise going about their daily life, each unique set of data is expressed in the architect's work.

Architectural theory and our current definitions of architectural style and influence are built upon this nebulous, genomic connectivity; yet the underlying structure is only ever exposed in miniscule chunks, often relinquishing much of the praise for the profession's stylistic development to chance or fate. With the rapid progression toward data as a solution for deeper understanding, the remapping of architectural history is not far away. Like the Human Genome Project (HGP), aimed at identifying and mapping all of the genes of the human genome¹, The Architect's Genome aims to map and identify the connections and influences that form the basis of the profession, thereby allowing a reexamining of current perceptions of style.

History favors the story of the lone architect; an image put forth since the transformation of the "master builder" into the tortured, creative genius whose vision was then executed during construction. Although architecture still has not recovered from the schism between the notion of the "designer" as a singular entity and the truth of the social and collaborative profession it remains, slowly we are seeing the return of the collective firm and the death of "Starchitecture." Social networks are an integral part of the historical analysis of people, but the large-scale dynamics of the design-world have often been overlooked in critical theory and history in favor of the individual-centric approach for categorizing style.

RELATED WORK [BIBLIOGRAPHY]

Scaled in Miles

<http://fathom.info/miles/>

A Fathom visualization which reveals the intricate web of session musicians who collaborated with Miles Davis to bring to life the grand opus of music associated with the Jazz giant. The style and concept for this visualization played a large role in influencing the form of our project, and many of our ideas came initially as improvements for this page.

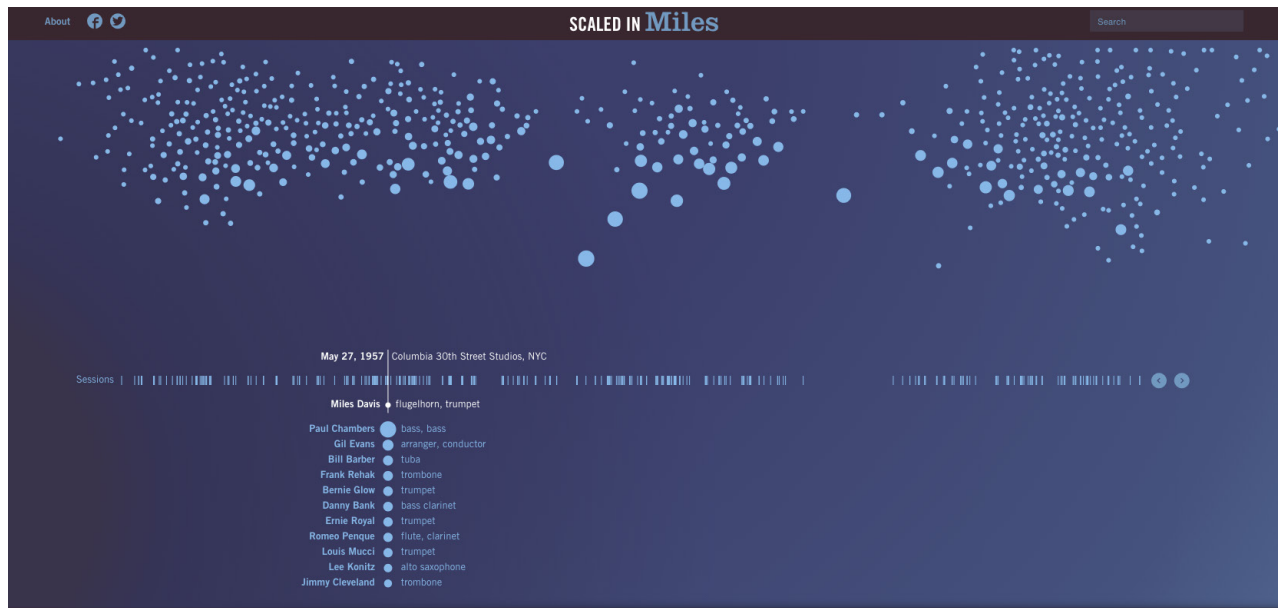


fig. 2 Scaled In Miles

Universeries

<http://universeries.hellohikimori.com/en/>

A visualization by Orange Cinema Series and an independent designer, HelloHikimori, Universeries is meant to show the complicated connections between Producers/Directors/Creators of US television series. Some of the data visualization elements were overlooked in this project in favor of creating a more aesthetically pleasing GUI. The three-dimensionality detracts somewhat from the usability, but is also quite pleasing and served as inspiration during the design process.

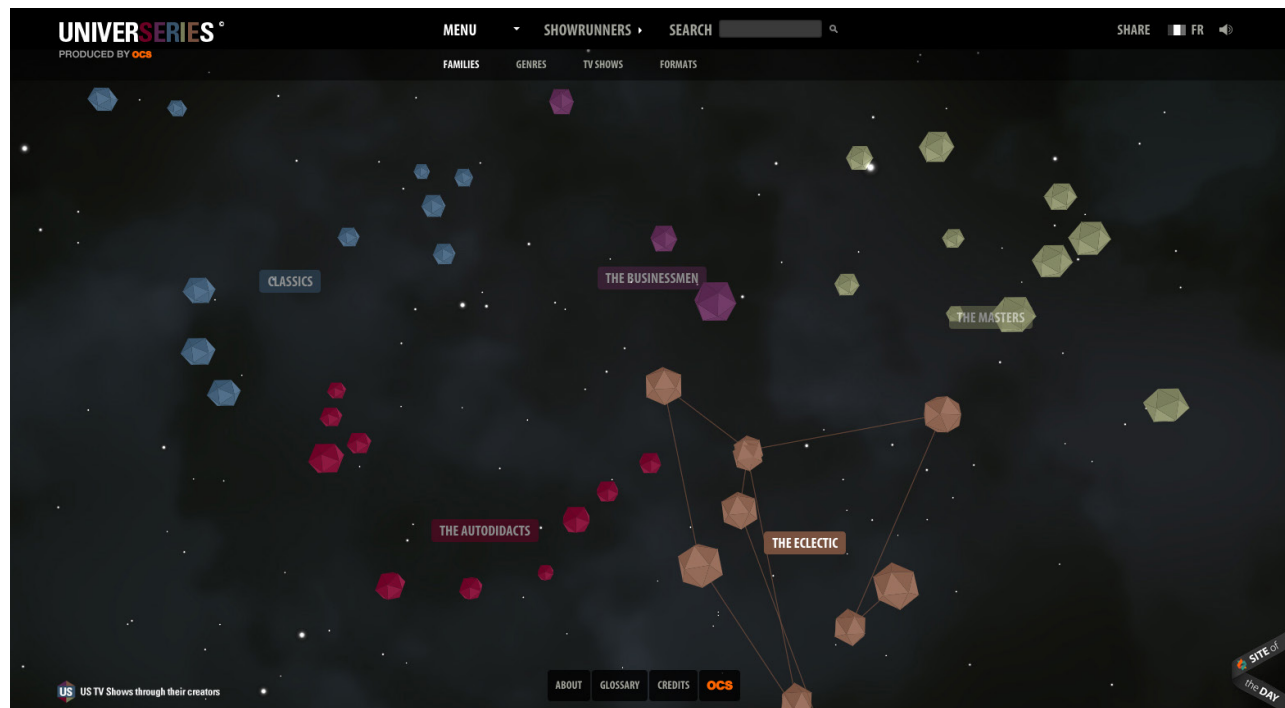


fig. 3 Universeries

Architect's Genome v.1 <http://jakerudin.com/123081/6281507/home/the-architects-genome>

The original study of the project, developed by Jake Rudin in conjunction with MDes thesis research at the GSD. The idea for comparing architects in a social network originally started by considering the way in which style and the profession as a whole develop over time. This visualization is a continuation and expansion of this idea.

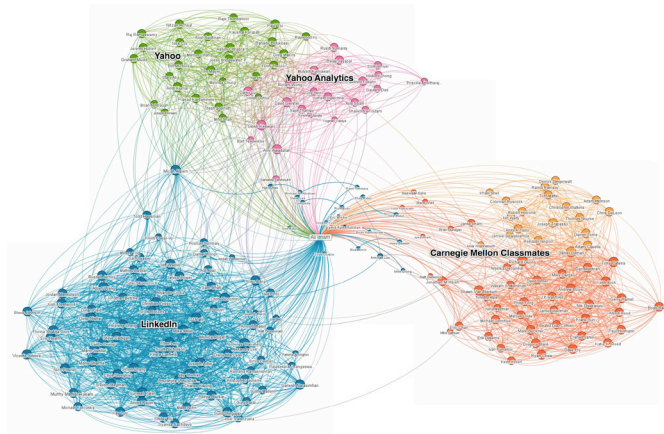
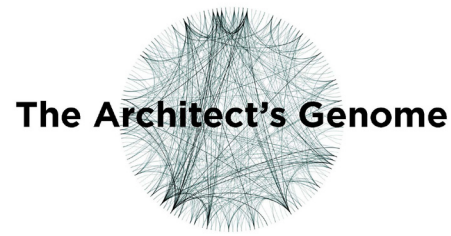


fig. 4 LinkedIn Maps



AZIZ BARBAR | AKSHAY GOYAL | JAKE RUDIN

fig. 5 The Architect's Genome v.1

Google Ripples <https://plus.google.com/ripple/details?url=>

A paired visualization with Google+ exploring the connections created between users of the application.

LinkedIn Maps <https://www.youtube.com/watch?v=PC99Nw2JX8w>

A paired visualization with LinkedIn, sadly no longer supported, exploring the connections between users.

Google+ ripples

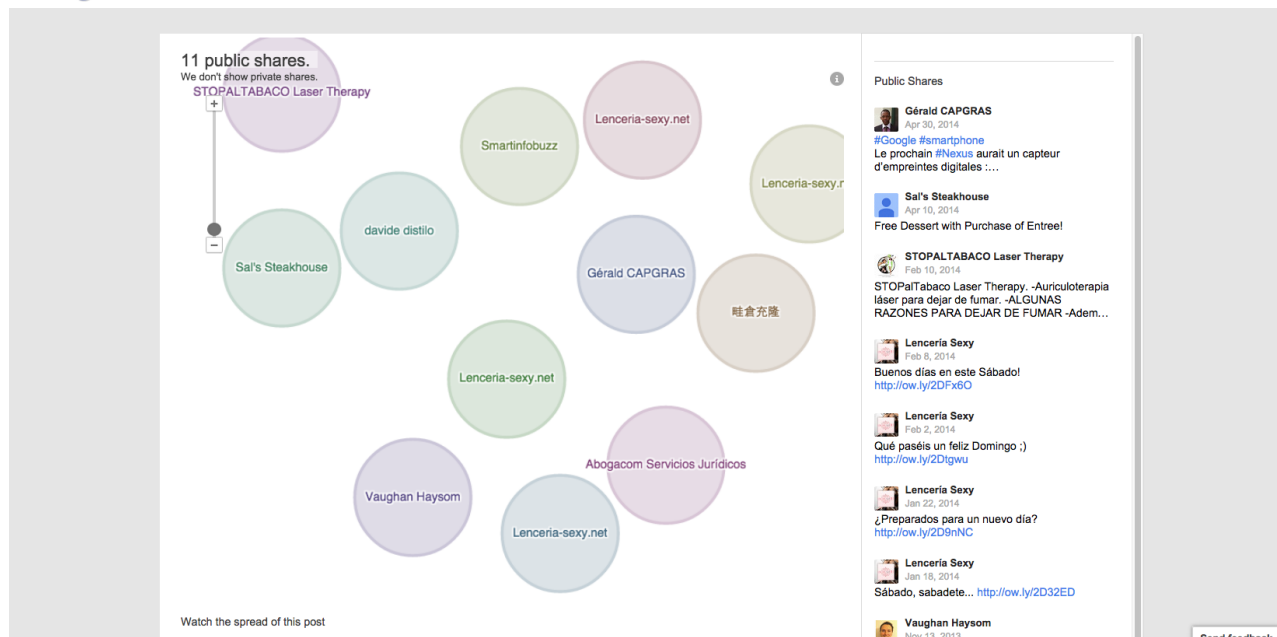


fig. 6 Google Ripples

QUESTIONS

The objective of this visualization is to address the user groups invested in the understanding of architecture and architectural history and answer questions that relate to the influence and change in the profession over time. The intended user groups of the project are architects, architectural theorists and historians, students of design and architecture, as well as passionate hobbyists.

The questions that might be answered by such a visualization include direct questions (i.e. What universities did Renzo Piano attend and what other famous architects were there

in the same years?) as well as larger indirect questions that include the conceptualization of style (i.e. What path do most architects take to become Pritzker Prize winners? or What is the extent of Le Corbusier's influence on other architects he has collaborated with and his connectedness over his lifespan?). The visualization is meant to reveal trends through statistical comparisons of architects' lives where previously they may have only been regarded for a particular style or genre of work.

DATA

The data for this project comes from a variety of sources, as no database such as this currently exists. Much of the data has already been compiled and there is a template which has been completed for all of the 39 Pritzker Prize winners (the "Nobel Prize" of architecture). These 39 were coded by hand and sourced from all over, but we have been looking into data scraping and are currently working with a list of 750 architects from Wikipedia as the main source, with a goal of having 300 architects with substantial enough data to use.

The scraping is being executed using a python wikipedia scraping library to query the links and pull the information from all of the "infoboxes" into a text file. That file can then be parsed into various data structures (including JSON) to create the final database.

However, much of the data is missing and one of our team members is spending a substantial amount of time and effort to continue filling in the missing data.

The data we are looking for has several parts depending on the organization of the visualization. Key information for forming each "object" (object meaning architect and their values and characteristics) include name, date of birth, date of death, university affiliations (attended and taught at), work affiliations (worked for, worked with, and firms operated), projects and buildings worked on, and time associations for all of these so that they can be related to a timeframe.

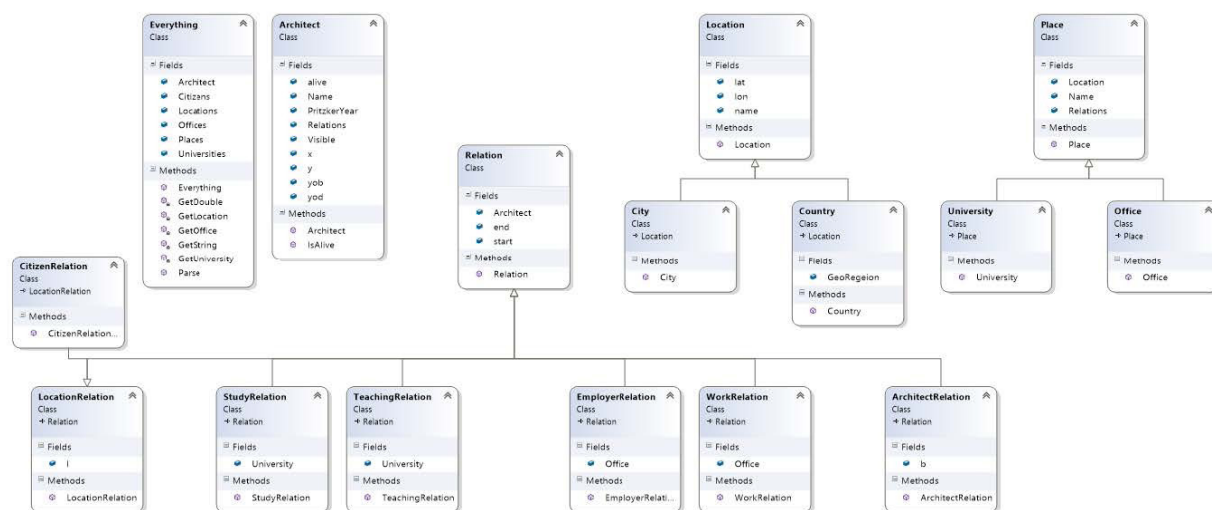


fig. 7 Data Class Structure

VISUALIZATION: INITIAL DESIGN

Initially, our ideas about the visualization ran towards traditional "social network" visualizations, using a graph or node-based visualization. The idea of using a more non-traditional "set visualization" is also appealing. The visualization needs to be able to represent a cross-section of time, as well as the connections existing in any given period, so the collapse and expansion of time is an interesting and important part of the visualization.

The other interesting component of the project is being able to switch between graph visualizations where on one the central nodes may represent architects with

their respective universities and workplaces configured around them, yet on another the universities may be the main nodes with the people attending them at a certain time shown as connected to them. These different “gravitational settings” will allow for different interpretations of the information.

MUST-HAVE FEATURES

Graph visualization of architects and their connections (links) showing university, work and location relationships. Must be able to scroll across the time span and view the change in the connections. Must be able to select any given architect and reveal a deeper level of information about their date of birth, place of birth, name, etc.

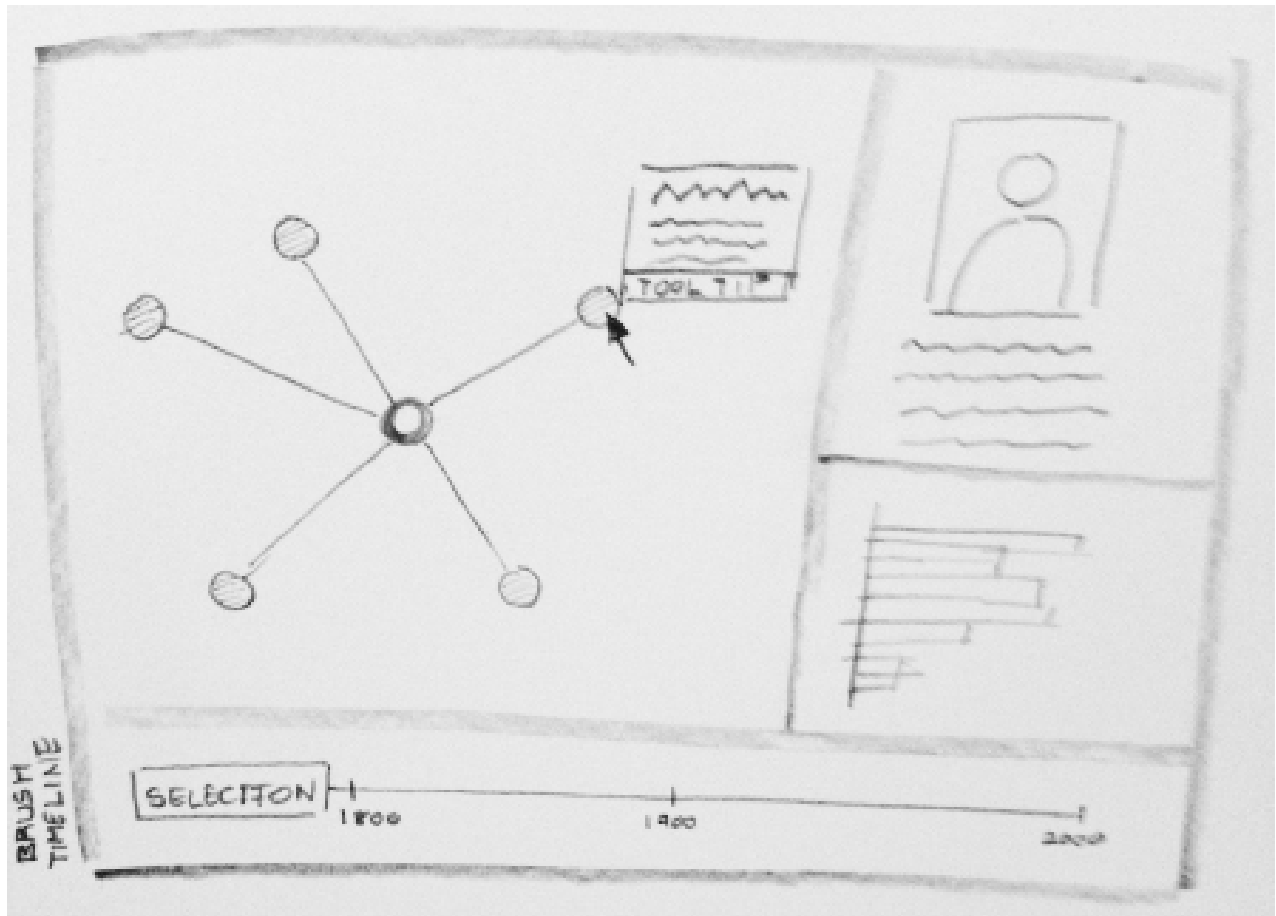


fig. 8 Initial Design Sketches

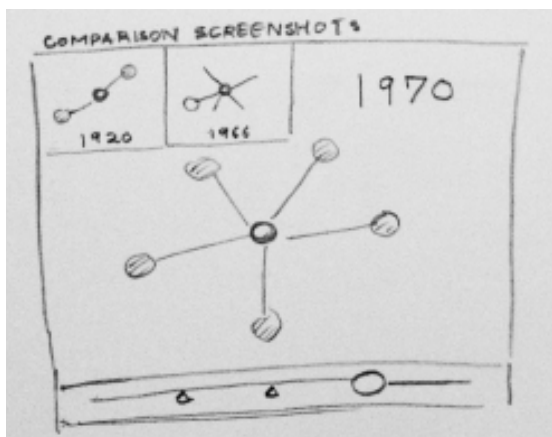


fig. 9 Zoom-in on relationships

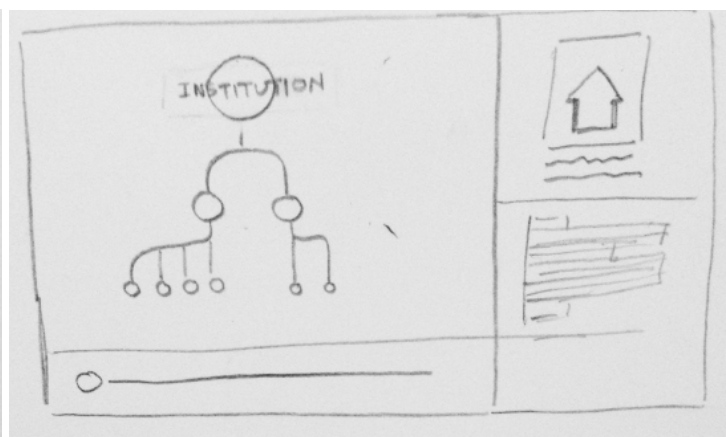


fig. 10 Shifting Focus of Tree Structure

DATA SCRAPING

Though as stated earlier our data required a lot of clean-up, we did have to create the entire database ourselves. We used scraperwiki and BeautifulSoup to search through a list of URL's and return a detailed categorization of pages where the infobox was labeled as "Infobox_Architect." From here we pulled the information from each box and dumped it directly into a new JSON file that would be the start of our database. The result was over 700 architects, but many of whom had unreliable information or gaps in their data. With interconnectivity being the main goal we needed a way of acquiring more data, from other sources that would be complete and have the social aspect.

```

1 import scraperwiki
2 import lxml.etree
3 from bs4 import BeautifulSoup, SoupStrainer
4 import wikipedia
5 import urllib, urllib2
6 import re
7 import json
8
9
10
11▼ def ParseTemplates(text):
12     "Extract all the templates/infoboxes from the text into a list"
13     res = { "templates":[ ], "categories":[ ], "images":[ ], "wikilinks":[ ], "flattext":[ ] }
14     templstack = [ ]
15▼    for tt in re.split("\{\{|\}\}|\{\{|\\}\\|\\[|\\]|\\)", text):
16        if tt in ["{{{", "{{", "[["]:
17            templstack.append([tt, [ ] ])
18▼        elif templstack and tt in ["}}}", "}}" , "]" ]:
19            templstack[-1][1][-1] = "".join(templstack[-1][1][-1])
20            templstack[-1].append(tt)

```

```

118 def GetWikipediaCategoryRecurse(categoryname):
119     "Downloads everything in a given category and all the subcategories"
120     prestack = [ categoryname ]
121     usedcategories = set()
122     result = [ ]
123     while prestack:
124         lcategoryname = prestack.pop()
125         if lcategoryname in usedcategories:
126             continue
127         for d in GetWikipediaCategory(lcategoryname):
128             if d["title"][:9] == "Category:":
129                 prestack.append(d["title"][9:])
130             else:
131                 result.append(d)
132         usedcategories.add(lcategoryname) # avoids infinite loops

```

```

139 infobox_architect = []
140
141
142 ▼ with open("architects.txt", "r") as ins:
143     urls = []
144     for line in ins:
145         urls.append(line.rstrip('\n'))
146
147 wiki = "http://en.wikipedia.org"
148 ▼ for i in urls:
149     goto = wiki + i
150     request = urllib2.Request(goto)
151     response = urllib2.urlopen(request)
152     soup = BeautifulSoup(response.read())
153
154     title = soup.find('h1', class_='firstHeading')
155     title = title.text
156
157     val = GetWikipediaPage(title)
158     res = ParseTemplates(val["text"])
159     persondata = dict(res["templates"]).get("Infobox architect")
160 ▼ if persondata:
161     infobox_architect.append(persondata)
162     print infobox_architect
163
164 json.dump(infobox_architect, open( "arcc.txt", "w" ) )
165

```

VISUALIZATION: MILESTONE DESIGN

The milestone implementation of our visualization is a very similar structure to the first, but with a more intuitive user interface. The visualization is sorted into 6 parts: the top div contains all of the architects bound to svg bubbles, the left most middle div is a profile of any selected architect, the second middle div is the architect's node, the third middle div contains the places, universities, and offices that the person is associated with, the fourth div contains all architects that have relationships to the selected architect, and the bottom is the year slider. The structure of this new visualization allows for all architects to be visible in frame, the organization of the top nodes is by year, earliest to the left and most current to the right. The central area however is where the connections are revealed. The pitfalls of this current visualization are that you are always “scaling” the visualization to an architect and never to a university or place. This means that while you can see who was connected to a university at a given year, it still has the architect in question as the focus. What will be a further step is to implement a system that allows brushing of the year to allow you to see all people related to a place or office during that period.

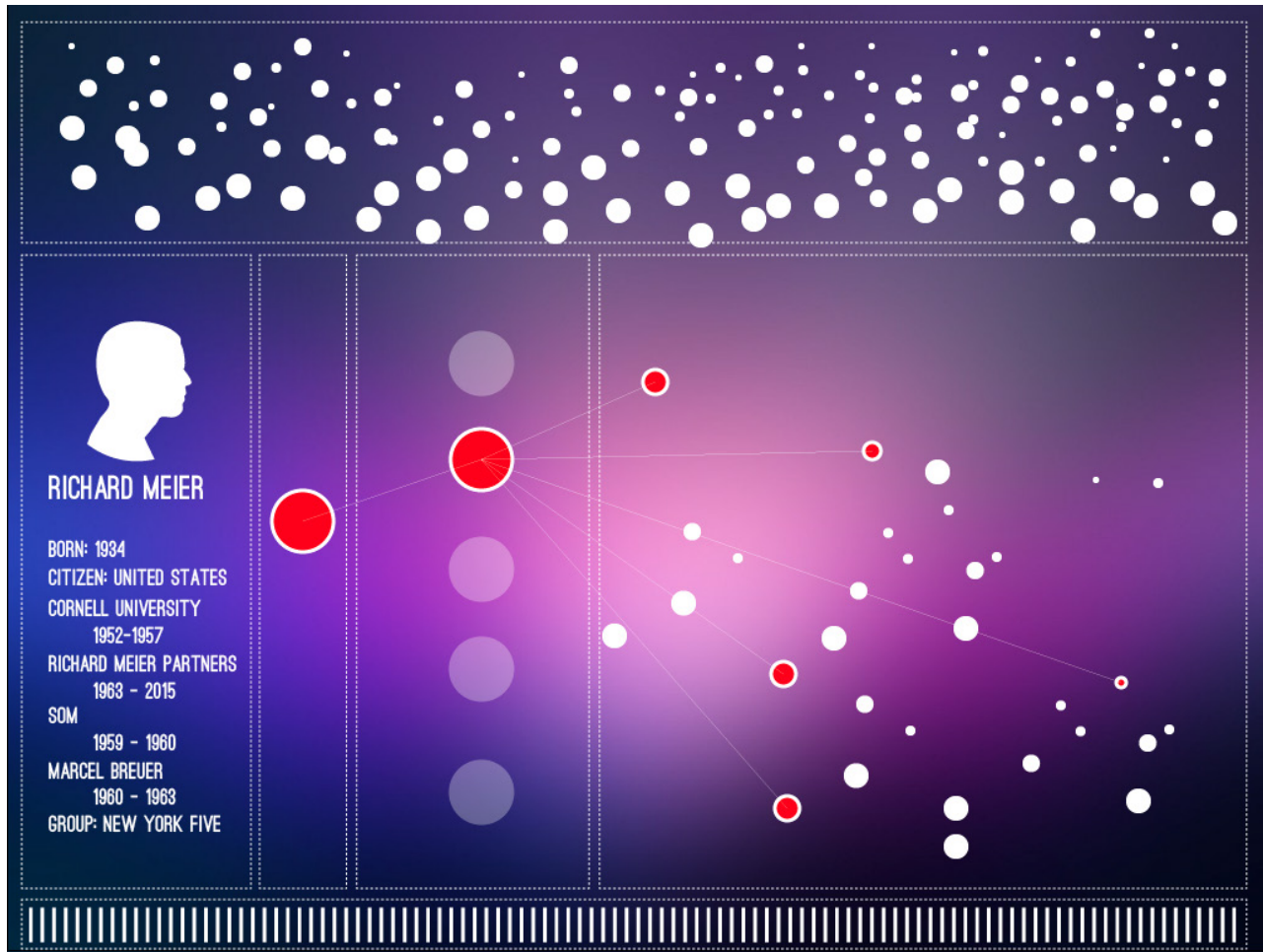


fig. 11 Current Design Mockup

Discussion of our milestone project with the professor led us to reconsider our representation and encoding of the data we are trying to portray. Initially apparent was the need to have some kind of differentiation between nodes of different classes: universities should not appear the same as people, and they should not appear visually similar to the offices or cities. Clarity would also be benefitted by having visual indicators of differences in people, which could be done by having the photos of individuals on top of their nodes. The lack of a second visualization also brought about changes for the final, noting the need for a timeline/barchart for the lower portion. Finally, the interactivity and on hover functionality needed to be used to its fullest.

The Architects' Genome

Search Build An Architect

The following visualization shows Architecture Relationships



Name: Paul Rudolph

Citizen: United States of America

Born: 1918

Death: 1997

Harvard University
1940-1947



The Architects' Genome

Rafael Search Build An Architect
rafael moneo shows Architecture Relationships



Name: Frank Gehry

Citizen: Canada

Born: 1929

University of Southern California
1951-1956
Harvard University
1956-1957

Gehry Partners
1962-2015



The Architects' Genome

Search Build An Architect

The following visualization shows Architecture Relationships



Name: Frank Gehry

Citizen: Canada

Born: 1929

University of Southern California
1951-1956
Harvard University
1956-1957

Gehry Partners
1962-2015



DESIGN EVOLUTION

The evolution of our design was not characterized by any wild swings but rather by gradual adjustment towards cleaner design. The initial design was inspired by the efficient layout of Homework 3, one webpage that was able to dynamically show the story it presented. Our milestone rendition reflects that view. However, as our pieces began working together, we began to grow into a new vision.

The first major change was the eradication of a bar graph showing the architect's top connections over their lifetime. Although it could have easily been added as a side graphic, our conversation with Professor Lex led us to think that there were more elegant ways with which to showcase proximity. With that in mind, we developed the "depth" variable to gauge the strength of connections and have presented the data accordingly in the main visualization. To enforce the importance of relationships, we made it so that the timeline at the bottom compares the lifespan of the selected architect and their closest pier, in tiers of proximity. We find this to be a more nuanced and elegant way of showcasing the information that the line/bar graph in our original sketches presented.

Secondly, our concept of what the timeline at the bottom should be has changed over time. While the milestone design called for rectangles that would resize dynamically for the selected architect with the capability to adapt on hover, our session with Professor Lex inspired us to create a new secondary visualization that would remain true to the focus of our project: connections. As such, the timeline is now more streamlined and colorful, but also full of information. The importance of color had been stressed throughout the course, but it was never as evident as when seeing the new timeline. Color coordinated across architects, the locations where linkage was built are attractive and evident to anyone using our website.

Another major change was the design and layout of the nodes. The milestone design called for a more textured background and plain, white nodes. Our original belief was that this would provide a sleek, minimalist visualization. However, we realized that this did not line up with an important tenet of our class and visualizations, the effective presentation of information. Without any built in differences, no one would understand our visualization. We decided to insert the pictures of the architects into the nodes and adapt their initial size according to their notoriety. Additionally, we assigned an array of colors to schools, offices, and cities. Further sizing and opacity choices were made in order to remain informational but not distracting. Tooltips with relevant information have also proved invaluable in their capacity to provide immediate feedback to the user.

Finally, our management of the depth variable was the biggest struggle throughout the project. Although invaluable for its classification of node strength, it was difficult to gauge how many links should be shown. In previous iterations, we presented a depth of three. While this showed impressive linkage through the smallest of architects, it provided a ludicrous number of nodes for the more famous ones. It took some time to find the right setting that would keep the small architects interesting while not rendering the bigger names unmanageable.

IMPLEMENTATION

The nodes come in two forms of interactivity, in the cloud and on drop-down. The first iteration is designed on a timeline, with birthdate going from left to right. As previously stated, the size of these nodes is decided based on an array that accumulates the number of universities and offices inhabited. Both of these variables are in place so that the user has information fed to them as soon as the nodes pop on screen. On hover, the node increases in size and is ringed by a shade of red that matches the design of the visualization. More importantly, a tooltip displays the name of the architect, their

birthdate, and their citizenship. When one moves away from this, the nodes shrink back to size. Ultimately, we tried to find the balance between clean design and information to create an environment that a user would be interested in interacting with.

The node's second form is deployed on click. First, the click triggers the architect info box. The box contains all the information anyone could hope to have about this architect, with a large image of them on top and a list of related architects at the bottom. If the architect has a large amount of information, the box is able to scroll to display. The box is also active on search. Nodes and links that have come down into the main arena display further information upon hover and can become the active party when clicked. The rest of the visualization adjusts accordingly. The position of the nodes is central to the visualization, with the svg segmented according to tiers of proximity. At a glance, the user will be able to see the connections and understand their importance.

Our timeline is also able to respond dynamically to the changed nodes. As the architect changes, timelines appear on top and bottom of a main timeline. When one hovers on a timeline, the office or university attended displays its name and the time period attended is shown. Additionally, we intend to have the rectangles that correspond to a node that has dropped down into the live visualization light up when hovered upon. That is to say, if the "Harvard University" node is hovered upon, all timelines with "Harvard University" would have that rectangle glow.

Finally, we have the search box and build an architect. The search box has an autocomplete function that allows curious users to search for their favorite architect, regardless of spelling capacity. Our "Build-an-Architect" page provides further interaction with the user, allowing them to improve the project and take ownership in it. This would hopefully expand our database while inspiring people to keep coming back. Since we do not know how to dynamically update our json file, this poses a prime opportunity for us to filter through submissions. Functionality is excellent in both these spheres and should improve the user experience.

EVALUATION

The most interesting part about using the visualization, from the view of an architecture student, is that the data reveals inconsistencies with traditional theory and the architectural canon. For example, the nodes which are often "best connected" are often not the most famous. The simple display and visual linkage of people and places allows for easy understanding of the relationships shown. The visualization works well and is a good representation of the work we had to put in to get the database up and running from data scraping through to implementation. The next steps to further the project would be to implement the server database to allow for constant updates via the user input page, to implement a tagging system to allow for new relationships to be created between existing architects (i.e. "father of", "partner of", "husband of", etc.), and to further develop the layout of the main visualization and the highlighting and on-hover functionality.

SPECIAL THANKS & COLLABORATORS

Our special thanks to all the collaborators and people who've helped us get this together.

Aziz Barbar | Collaborator on Version 1

Akshay Goyal | Collaborator on Version 1

Alexander Lex | CS 171 Advisor and Design Critic

Christopher Morse | Extensive Data Scraping Assistance

Panagiotis Michalatos | Conceptual Guidance and Version 1 Advisor