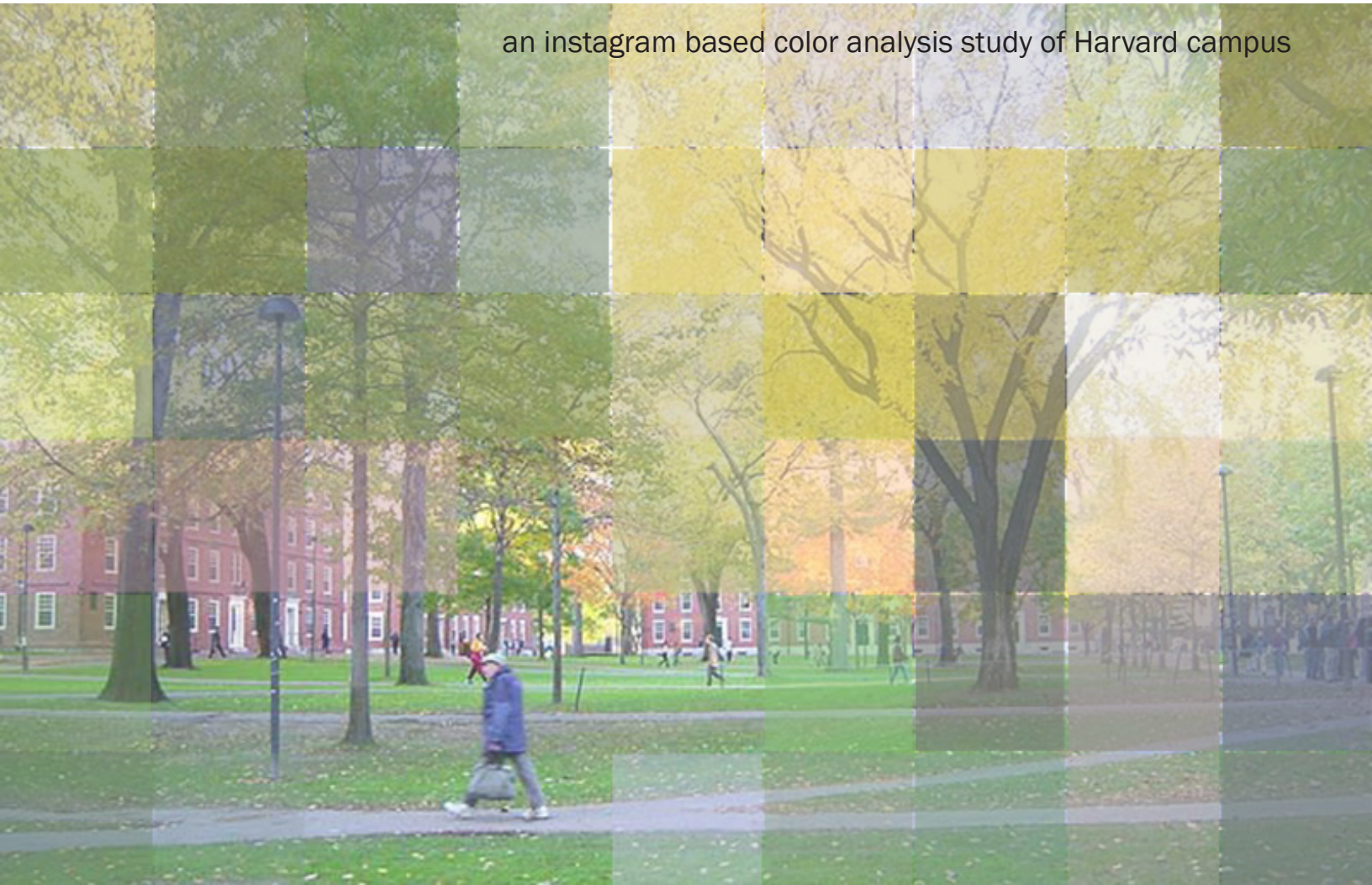


an instagram based color analysis study of Harvard campus



— LANDSCAPE THROUGH COLOR —

Alkistis Mavroeidi - Jonathan Taratuta-Titus - Steven Vasilakos
developed for the course *CS171 - Visualization* (Harvard SEAS)
Spring 2015

Instructor: Alexander Lex
Teaching Fellow: Jordan Dominguez

– LANDSCAPE THROUGH COLOR –

Alkistis Mavroeiði - Jonathan Taratuta-Titus - Steven Vasilakos

OVERVIEW

Landscape through color is a project that analyzes Instagram pictures taken around the general area of Harvard Yard and then extracts and visualizes the primary colors. The visualization is meant as a tool of landscape exploration, as well as a way of analyzing the way people perceive their environment and how seasons or events affect this relationship.

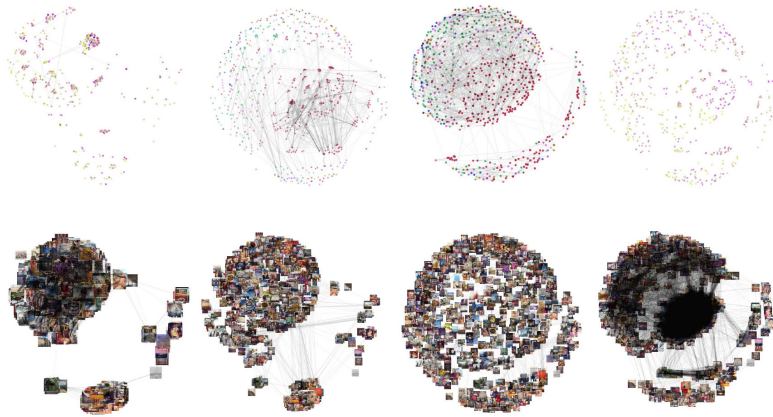
BACKGROUND AND MOTIVATION.

Starting a new academic program usually implies a big change on students' lives. Many move from distant countries and different cultural backgrounds and have grown up surrounded by completely different landscapes and visual experiences. Through discussions with international students who moved to Cambridge from a diverse range of backgrounds, and by trying to understand the visual implication of our life's changes, we became very interested in the idea of studying colors from pictures we have taken during the past months of our lives. By browsing our files we could clearly see the change, but we wondered if there was a way to computationally analyze and visualize the results of a study that would include a wider range of pictures and people. Our motivation was to prove that there is a visible alteration of colors signifying that transition (from our home countries to Boston) and to try to answer the question of how the change of location or the change of seasons affects the way we perceive and document our environment.

In order to test that idea, during the Fall semester we set off to create an application that performs a color analysis on pictures. The project was completed for the course *Introduction to Computational Design* at the Graduate School of Design (developed by Kritika Dhanda, Qurat-ul-ain Malick, Alkistis Mavroeiði). For the data collection, we decided to work with a social network to extract and study information. Instagram proved ideal for this application since it is a picture-oriented social network and because of the type of information we can extract from it (geolocation of the uploaded picture, timestamp etc.). Using its API, we developed a color analysis script, that extrudes and maps the primary colors of Instagram pictures, who are tracked and downloaded based on time and geolocation. Even though the computational part of the application proved successful, we believe that there is so much more that one can extract from that data, and a strong visual representation is key to help us explore the possibilities of a color based analysis of landscape. Only through a visual representation can this data come to life and reveal clusters, color patterns, peaks and other parameters that can reveal human behavior in response to landscape and landscape change. And this is where CS171 came in.

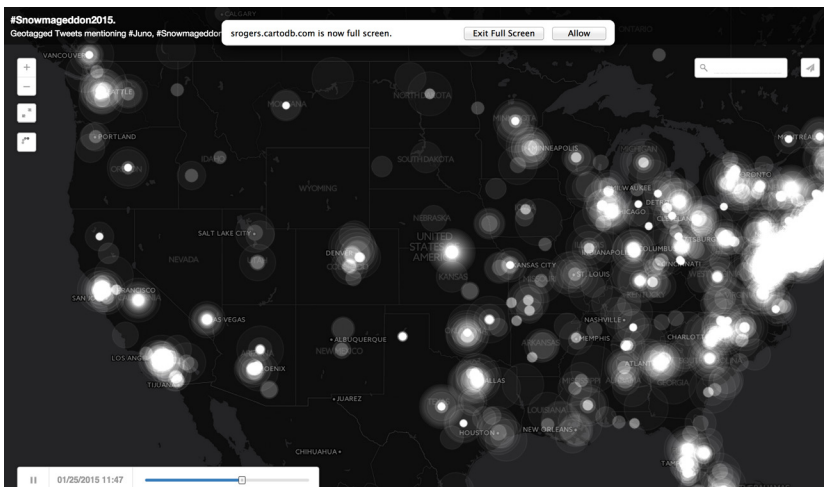
For the CS171 Visualization project, we formed a team passionate about discovering the best possible way of exploiting the color data. By choosing the Harvard Campus as our case study, we decided to take this research further and aim to create a visualization that would cover a year's length of color analysis around the area. This way we can gain a better understanding of what conclusions can be drawn and we can set the ground for an elegant tool for future research.

RELATED WORK.



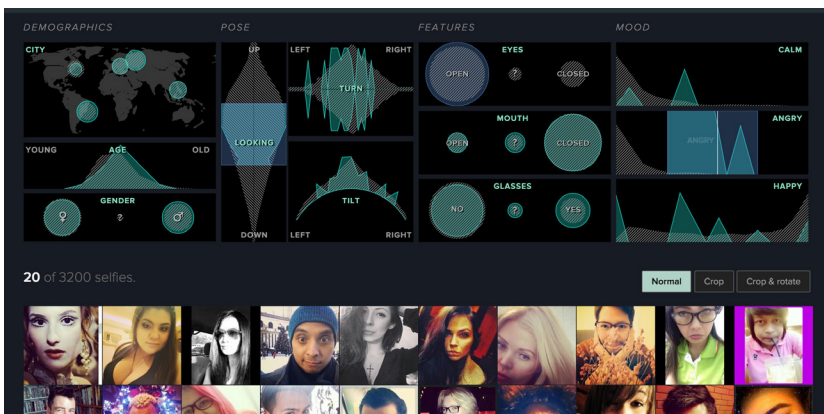
Reevaluating Notions of Proximity mapping space through Instagram

*Final Project for 6338 - Introduction
to Computational Design | Juan Pablo
Ugarte Michael Leef*



#Snowmageddon2015

*Geotagged Tweets mentioning #Juno,
#Snowmageddon2015, #Juno2015,
#BlizzardOf2015, #blizzard | produced
by Twitter's Data Editor Simon Rogers*



Selfiecity

*Investigating the style of self-portraits
(selfies) in five cities across the world |
Dr. Lev Manovich*

QUESTIONS.

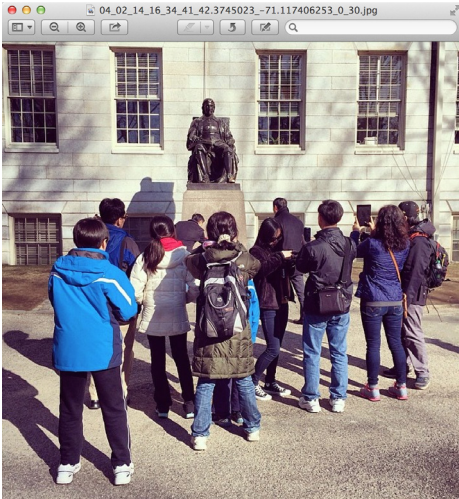
Our primary goal is the study of landscape through color. We are aiming to recreate the Harvard area landscape through the eyes of the Instagram users, primarily by using the way color is represented in the pictures. This mapping would help us answer a number of questions that either concern the visitors behavior towards the landscape or the landscape itself.

- What do primary colors represent in each location?
- How do seasons affect the colors of the landscape?
- Can we recreate a map solely based on color analysis of pictures?
- What do people choose to take pictures of in specific locations and how does that change with season?

DATA.

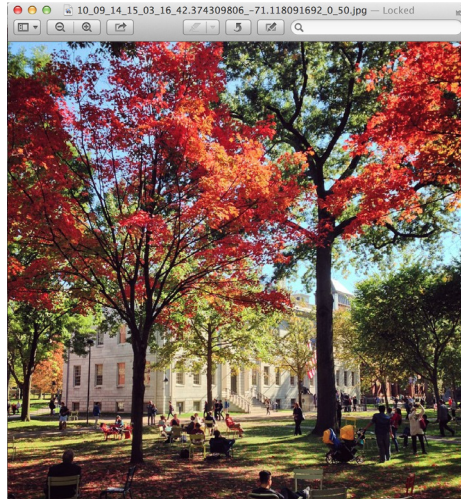
Though the previously conducted research, the data collection algorithm was almost entirely developed. Using C#, the script connects to the Instagram API, authenticates the developer's id and then performs a search based on the criteria we selected and downloads the pictures it finds. We chose to download pictures based on geolocation and time, including parameters like radius and number of downloaded media. The pictures are saved in a local file, storing all the relevant information in their name and thus available for the mapping later on.

Sample of downloaded pictures



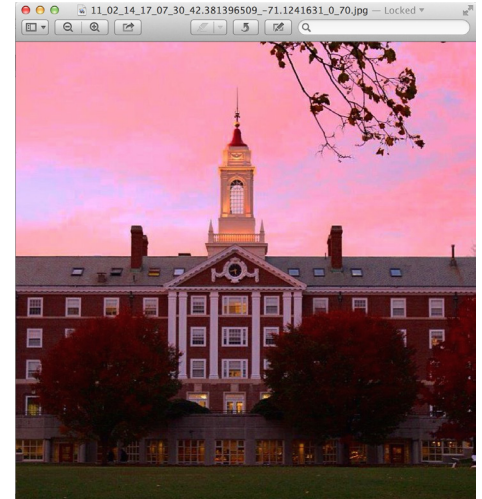
file name:
04_02_14_16_34_41_42.3745023_-
71.117406253_0_30.jpg

Date: 04/02/14
Time: 16:34:41
latitude: 42.3745023
longitude: -71.117406253



file name:
10_09_14_15_03_16_42.374309806_-
71.118091692_0_50.jpg

Date: 10/09/14
Time: 15:03:16
latitude: 42.374309806
longitude: -71.118091692

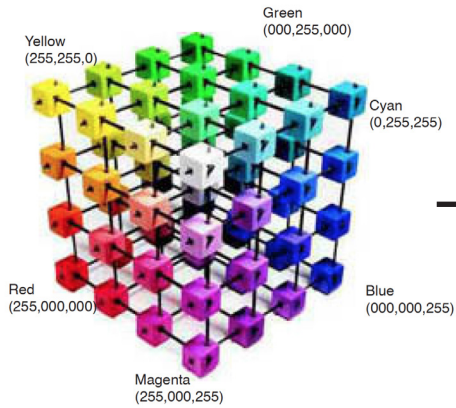


file name:
11_02_14_17_07_30_42.381396509_-
71.1241631_0_70.jpg

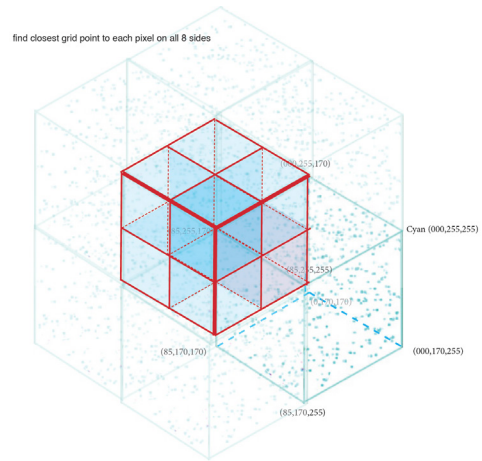
Date: 11/02/14
Time: 17:07:30
latitude: 42.381396509
longitude: -71.1241631

After the pictures are downloaded, carrying the date, time, longitude and latitude information in their titles, they are imported again into the program to be analyzed. Extracting the primary color of a picture is a computationally complex problem, but we decided to use a three dimensional mapping of each pixel's color and then average the most populated area. In more detail, the rgb values of each color are used as x, y and z coordinates in a 3 dimensional space. For every picture we analyze the color of each pixel and map it on the grid. We then divide it into 64 color bins (by creating a 4x4 division) and sort them from the most to the least populated one. By extracting the average of the color bins we get a pretty good approximation of the primary colors of each picture in their relative order. The testing stages were first developed in the Grasshopper plugin for Rhinoceros but then the algorithm was integrated into Visual Studio, and the colors were mapped according to geolocation and filtered according to time. With small alterations, we adapted it the script so that instead of mapping the colors, the data was exported to a json file for our use in this project.

Computational Color Mapping Method

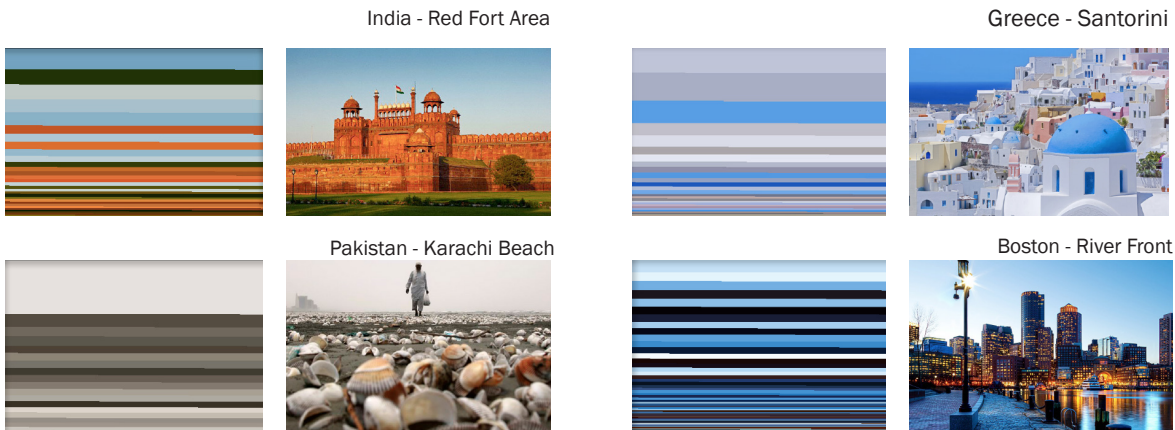


3-dimensional RGB grid



64 color bins

Color Analysis Tests from Grasshopper



RESEARCH COMPLETED PRIOR TO THIS CLASS

“Pan’s Colors”: Instagram Color Analysis Application developed for the course *Introduction to Computational Design*

Harvard University, Graduate School of Design
(Fall 2014)

Students: Alkistis Mavroeidi, Kritika Dhanda, Qurat-ul-ain Malick
Professor: Panagiotis Michalatos

DATA PROCESSING.

The data collection and data processing, despite the already developed script, was a copious process. Due to the limitations of the API access to Instagram, we could only download a limited number of pictures per request. Given that we wanted our dataset to be as complete as possible for a whole year's range, we performed the request for each day separately (365 times).

The process downloaded on average 75 pictures per day. However, Instagram is a widely used social network where people upload more than just landscape picture. Anything from quotes to old & unrelated pictures had to be filtered out. The filtering process was done manually and was aiming to leave in the dataset only pictures that represented how people perceive and document their outdoors experience. Filtered pictures, pictures focusing on individual objects, nighttime pictures etc. were kept as long as they were taken outdoors, however a wide range of other themes was rejected and deleted.

Pictures that were deleted were:

- _indoors pictures
- _quotes, cartoons, paintings and digitally produced images
- _pictures that seemed to be from different locations
- _old pictures

DATASET STATISTICS

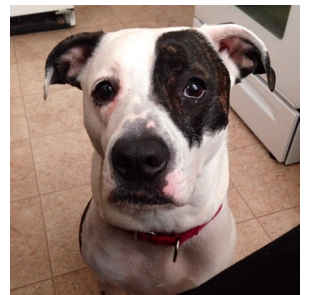
Total Number of Downloaded Pictures:	30,885	Total Number of Filtered Pictures:	11,063
Apr, 2014:	2,777	Apr, 2014:	721 (26%)
May, 2014:	2,683	May, 2014:	767 (29%)
Jun, 2014:	2,542	Jun, 2014:	857 (34%)
Jul, 2014:	2,801	Jul, 2014:	1,124 (40%)
Aug, 2014:	2,989	Aug, 2014:	1,242 (42%)
Sep, 2014:	2,667	Sep, 2014:	1,009 (38%)
Oct, 2014:	2,964	Oct, 2014:	1,189 (40%)
Nov, 2014:	2,627	Nov, 2014:	899 (34%)
Dec., 2014	2,883	Dec., 2014	773 (27%)
Jan, 2015:	1,184	Jan, 2015:	780 (66%)
Feb, 2015:	2,153	Feb, 2015:	926 (43%)
Mar, 2015:	3,115	Mar, 2015:	776 (25%)

SAMPLE OF PICTURES ACCEPTED:

- (1) Buildings from outdoors (2) Outdoor Selfies (3) Nighttime Pictures (4) Filtered Pictures (5) Pictures focusing on individual objects (6) Quotes on walls, Graffiti (7) Outdoors Food Pictures (8) Pets or Animals

SAMPLE OF PICTURES REJECTED:

- (1) Indoor Pictures of Building (2) Indoor Selfies or Pictures (3) Nighttime Indoors Pictures (4) Filtered Pictures with no clear location (5) Objects, Utensils or Other indoor equipment (6) Motivational Pictures or Quotes (7) Indoors Food Pictures (8) Pets or Animals from indoors (9) Old photos (10 -11) Pictures from other locations (12) Products or Unrelated Subject Pictures



After the pictures were downloaded and filtered, they were reimported to the program to analyze their color. We decided to extrude the 10 primary colors of each picture and save them in a json file, along with the longitude, latitude and time of each picture. We also used the file name as a unique id that would help identify the pictures later on (data.json).

After the json file was created, it was processed again in order to optimize our data information. Our script "weather.js" enabled us to both add weather information and reprocess and optimize the JSON file, and was run using node.js. The result was a dataset with reduced size of the RGB values and additional metadata. In particular, we added **HSV** (hue-saturation-value) values to each color that would later be useful in the color sorting process and weather data corresponding to the geolocation and date of each picture (output.json).

Sample of Picture with its data and the 10 primary colors



```
{ "id": "04_18_14_16_15_43_42.37088_-71.114113833_0_70.jpg",  
  "lng": -71.114113833, "lat": 42.37088, "date": 1397837743,  
  "color0R": 144.442855947333, "color0G": 123.550867551886, "color0B": 111.442060923901,  
  "color1R": 210.381303095037, "color1G": 152.346490138606, "color1B": 135.152205772766,  
  "color2R": 24.6754699083897, "color2G": 22.7496942163201, "color2B": 20.9170514964679,  
  "color3R": 236.177507409141, "color3G": 235.659625197763, "color3B": 226.595895447557,  
  "color4R": 99.1382998006163, "color4G": 74.6296296296296, "color4B": 54.2996193583469,  
  "color5R": 85.8003720663995, "color5G": 49.4250143102461, "color5B": 38.0698340011448,  
  "color6R": 221.237428243398, "color6G": 203.690088021431, "color6B": 178.338844240337,  
  "color7R": 175.412085718089, "color7G": 201.271396246506, "color7B": 203.052442433116,  
  "color8R": 176.705027494108, "color8G": 197.296936370778, "color8B": 178.245483110762,  
  "color9R": 49.8415257569799, "color9G": 78.7145104207629, "color9B": 88.5749115218246}
```



```

data.json
1  [
2  {
3    "id" : "04_01_14_13_45_30_42.371611111_-71.117361111_0_87.jpg",
4    "lng" : -71.117361111,
5    "lat" : 42.371611111,
6    "date" : 1396359930,
7    "color0R" : 149.91723723168, "color0G" : 133.594911176906, "color0B" : 126.76042746114,
8    "color1R" : 210.289021245933, "color1G" : 154.659389531283, "color1B" : 130.909792044316,
9    "color2R" : 43.7802153896446, "color2G" : 16.8227848101266, "color2B" : 38.2020776201805,
10   "color3R" : 210.501549615363, "color3G" : 206.168935746306, "color3B" : 158.844512701312,
11   "color4R" : 141.582361870067, "color4G" : 220.582999392835, "color4B" : 240.156162720097,
12   "color5R" : 78.0967287200962, "color5G" : 32.3141921306853, "color5B" : 45.741904722418,
13   "color6R" : 226.663782816229, "color6G" : 233.355395499489, "color6B" : 214.876235935902,
14   "color7R" : 81.8180245489744, "color7G" : 48.8611912019112, "color7B" : 74.3114754098361,
15   "color8R" : 44.5818997756171, "color8G" : 43.6950885066068, "color8B" : 83.7853403141361,
16   "color9R" : 117.408786210199, "color9G" : 169.748144601389, "color9B" : 212.159205171175
17   },
18
19   {
20     "id" : "04_01_14_14_21_06_42.375521667_-71.1163945_0_82.jpg",
21     "lng" : -71.1163945,
22     "lat" : 42.375521667,
23     "date" : 1396362066,
24     "color0R" : 75.8370661498152, "color0G" : 48.8891550913463, "color0B" : 80.4826998313054,
25     "color1R" : 116.675649705658, "color1G" : 98.2679449089363, "color1B" : 115.083078384378,
26     "color2R" : 234.655969043543, "color2G" : 231.067192197592, "color2B" : 211.809845073867,
27     "color3R" : 41.1182160959327, "color3G" : 19.62988831746817, "color3B" : 50.292387810607,
28     "color4R" : 223.903073913988, "color4G" : 212.73353777992, "color4B" : 175.76640185449,
29     "color5R" : 57.7516940719329, "color5G" : 39.5005923328437, "color5B" : 73.0770506563048,
30     "color6R" : 199.01540436457, "color6G" : 182.242190842961, "color6B" : 161.531022678648,
31     "color7R" : 71.703530259366, "color7G" : 41.2957492795389, "color7B" : 58.3393371757925,
32     "color8R" : 201.7, "color8G" : 183.825581395349, "color8B" : 197.225581395349,
33     "color9R" : 168.615776081425, "color9G" : 172.64631043257, "color9B" : 196.104325699746
34   },
35 }
36 {
37   "id" : "04_01_14_14_29_03_42.374309806_-71.118091692_0_78.jpg",
38   "lng" : -71.118091692,
39   "lat" : 42.374309806,
40   "date" : 1396362543

```

```

output.json
1  [
2  {
3    "id": "04_01_14_13_45_30_42.371611111_-71.117361111_0_87.jpg",
4    "lng": -71.117361111,
5    "lat": 42.371611111,
6    "date": 1396359930,
7    "colors":
8    [
9      {"r":150,"g":134,"b":127,"v":0.588,"s":0.153,"h":18.261},
10     {"r":210,"g":155,"b":131,"v":0.824,"s":0.376,"h":18.228},
11     {"r":44,"g":17,"b":38,"v":0.173,"s":0.614,"h":313.333},
12     {"r":211,"g":206,"b":159,"v":0.827,"s":0.246,"h":54.231},
13     {"r":142,"g":221,"b":240,"v":0.941,"s":0.408,"h":191.633},
14     {"r":78,"g":32,"b":46,"v":0.306,"s":0.59,"h":341.739},
15     {"r":227,"g":233,"b":215,"v":0.914,"s":0.077,"h":80},
16     {"r":82,"g":49,"b":74,"v":0.322,"s":0.402,"h":314.545},
17     {"r":45,"g":44,"b":84,"v":0.329,"s":0.476,"h":241.5},
18     {"r":117,"g":170,"b":212,"v":0.831,"s":0.448,"h":206.526}
19   ],
20   "uid": "p0", "weather": "clear-day"
21 },
22
23 {
24   "id": "04_01_14_14_21_06_42.375521667_-71.1163945_0_82.jpg",
25   "lng": -71.1163945,
26   "lat": 42.375521667,
27   "date": 1396362066,
28   "colors":
29   [
30     {"r":76,"g":49,"b":80,"v":0.314,"s":0.387,"h":292.258},
31     {"r":117,"g":98,"b":115,"v":0.459,"s":0.162,"h":306.316},
32     {"r":235,"g":231,"b":212,"v":0.922,"s":0.098,"h":49.565},
33     {"r":41,"g":20,"b":50,"v":0.196,"s":0.6,"h":282},
34     {"r":224,"g":213,"b":176,"v":0.878,"s":0.214,"h":46.25},
35     {"r":58,"g":40,"b":73,"v":0.286,"s":0.452,"h":272.727},
36     {"r":199,"g":182,"b":162,"v":0.78,"s":0.186,"h":32.432},
37     {"r":72,"g":41,"b":58,"v":0.282,"s":0.431,"h":327.097},
38     {"r":202,"g":184,"b":197,"v":0.792,"s":0.089,"h":316.667},
39     {"r":169,"g":173,"b":196,"v":0.769,"s":0.138,"h":231.111}
40   ],
41   "uid": "p1", "weather": "clear-day"
42 }

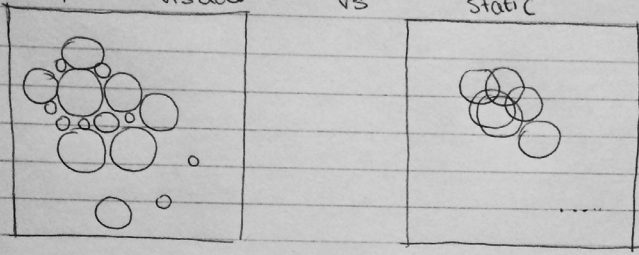
```

DESIGN EVOLUTION

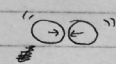
First thoughts on mapping and recreating the landscape

Map Visualization Sketches

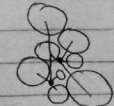
"dynamic" visual vs static



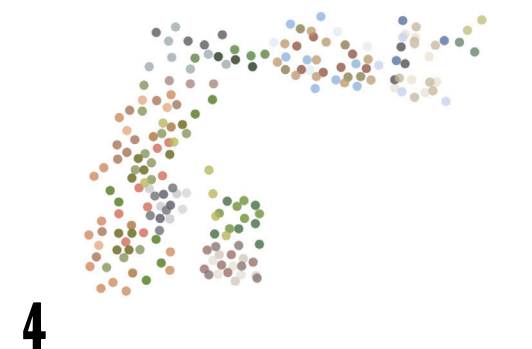
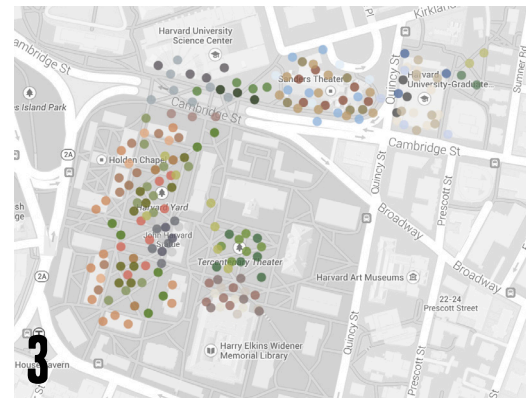
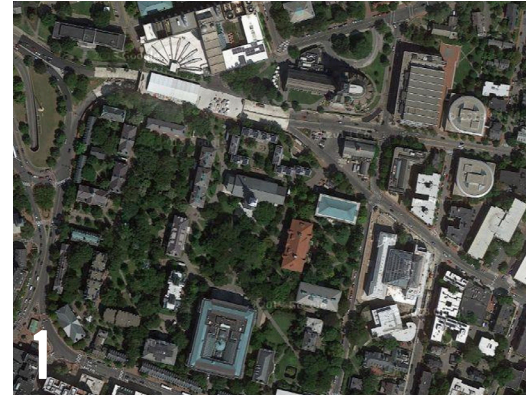
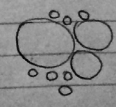
① create "colliding" nodes
"→←" that bump into each other instead of crossing each other without impact



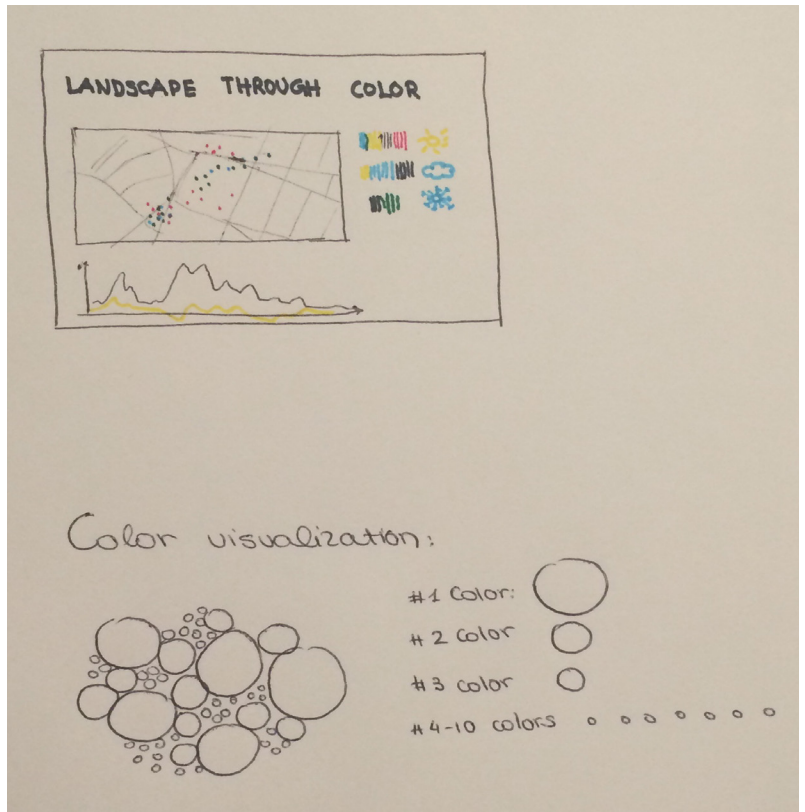
② create links between each node with its geolocation
so the force attracts them to the geolocation but the "pack" around it instead of overlaying one another.



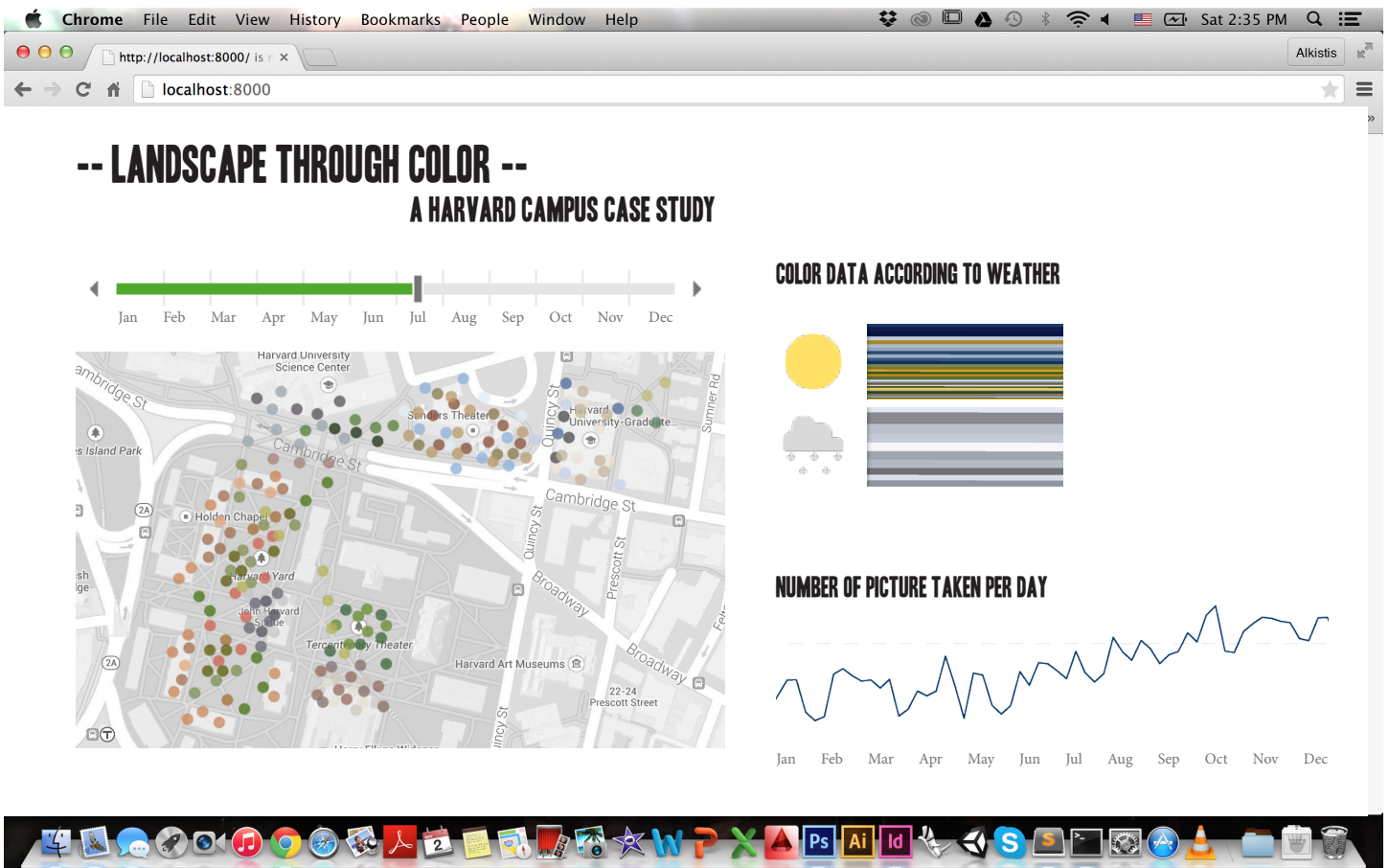
③ have 3 main colors and the 7 rest very small. so that the visualization becomes more populated.



Dynamic Node mapping ideas

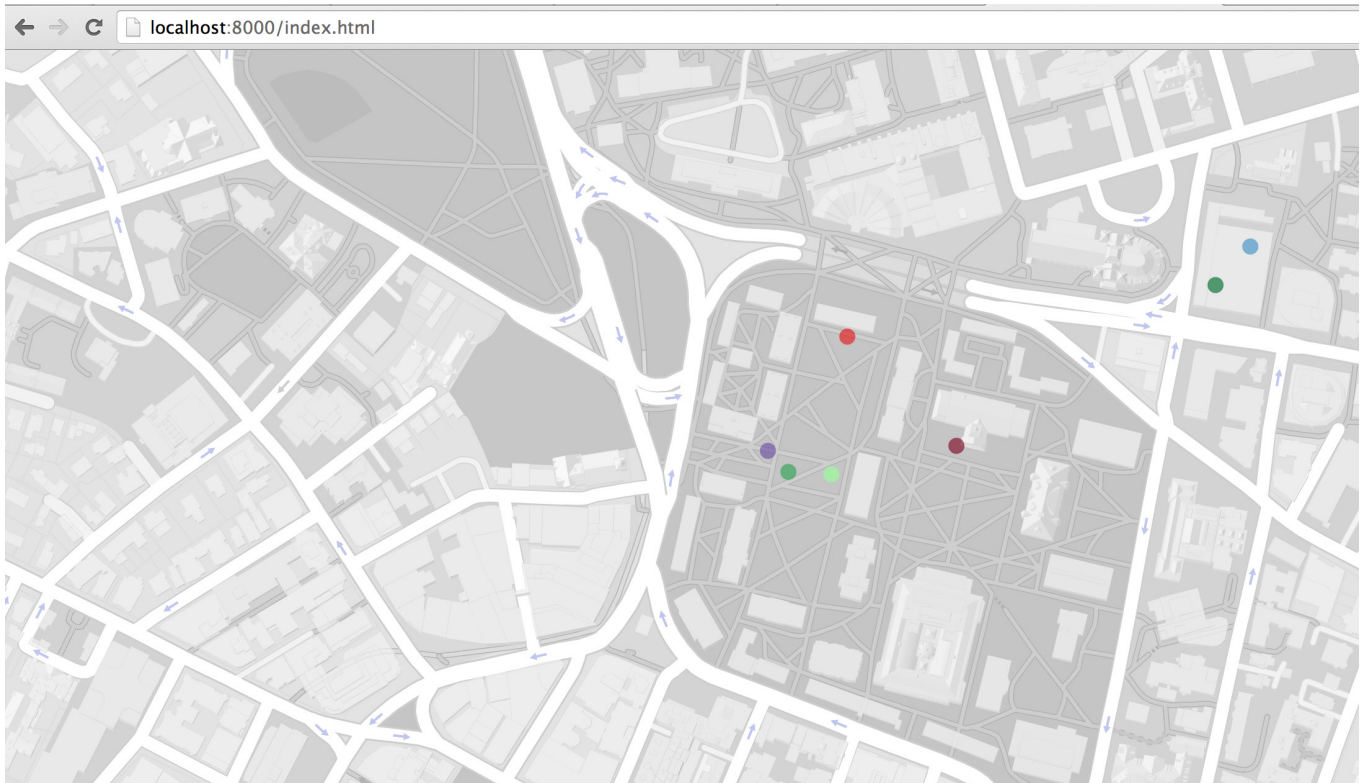


Layout and visualization ideas



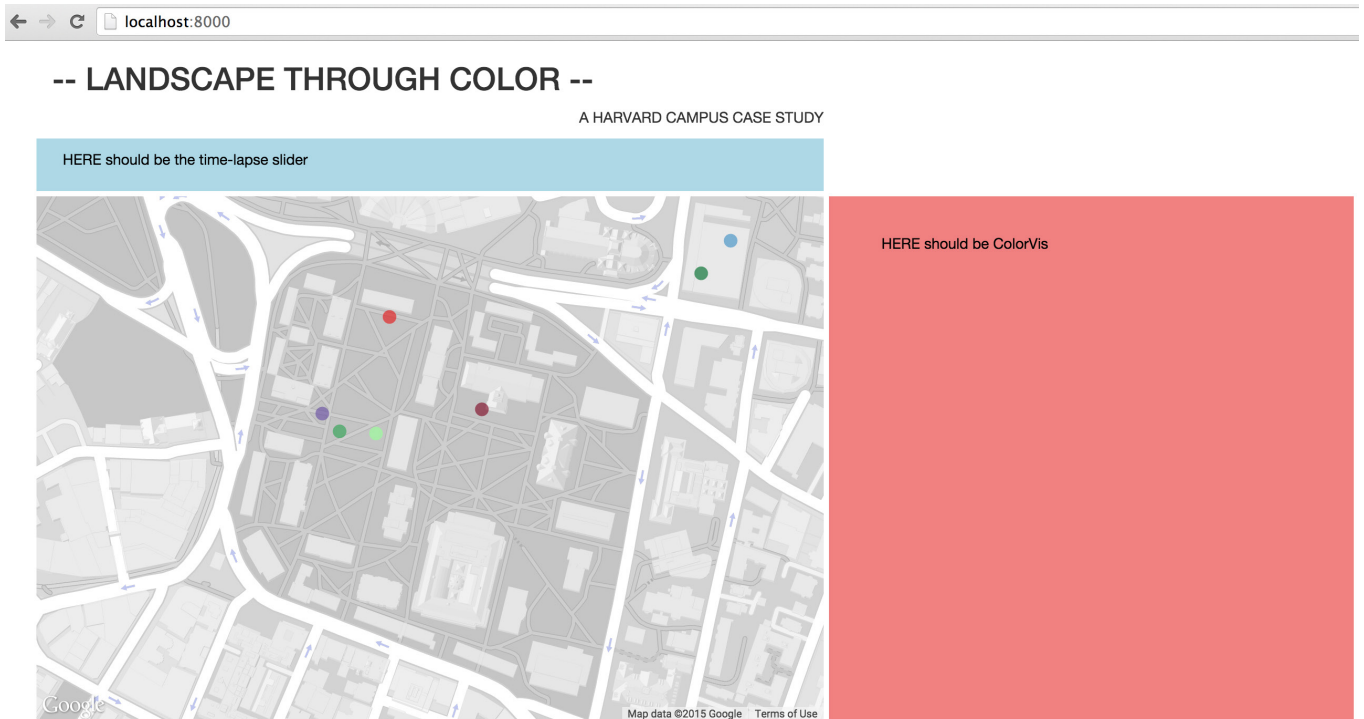
Layout organization ideas

IMPLEMENTATION PROGRESS



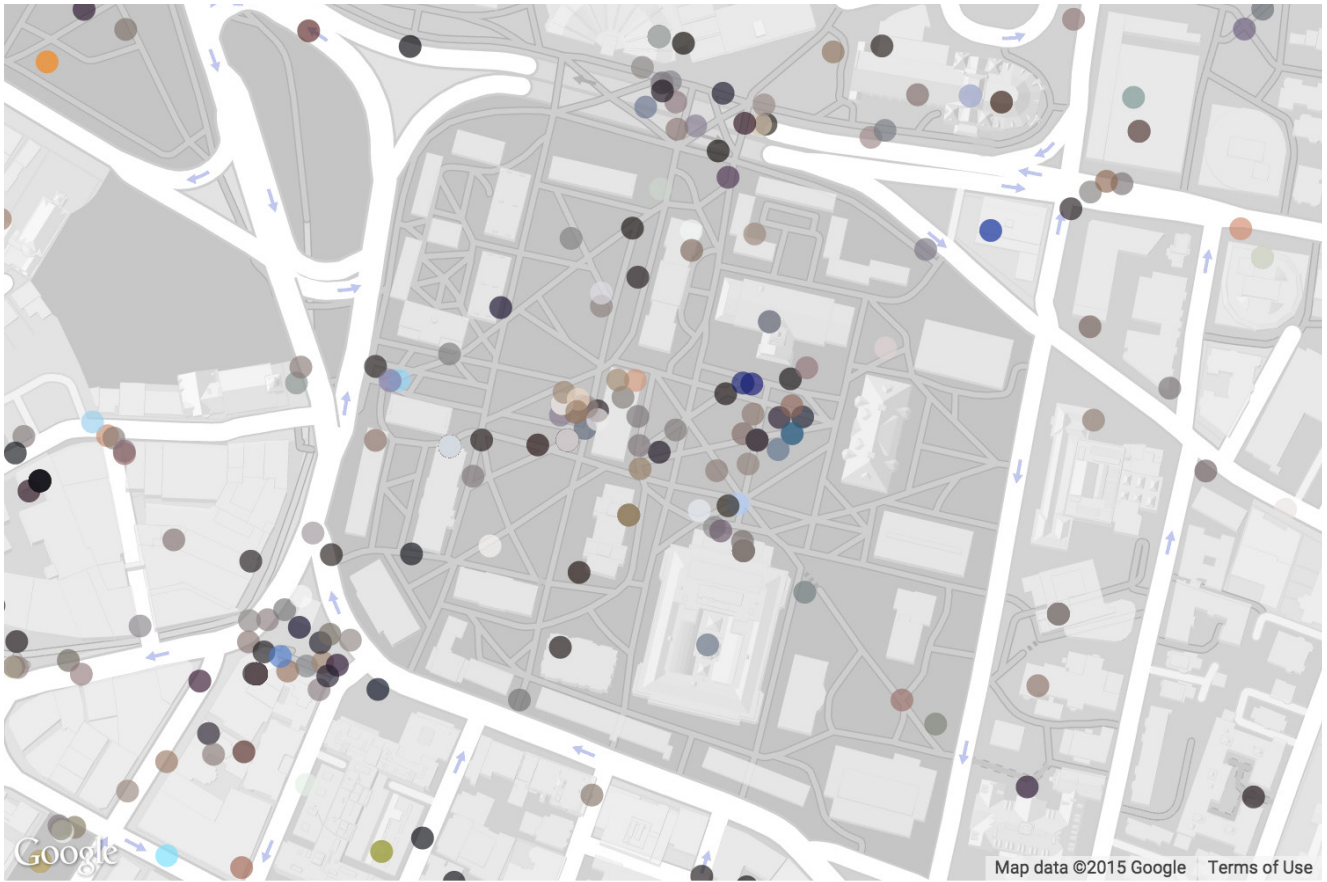
FIRST STEPS

Applying Google Maps API to project colors and first attempts to customize map visualization
Mapping of sample color data according to geolocation



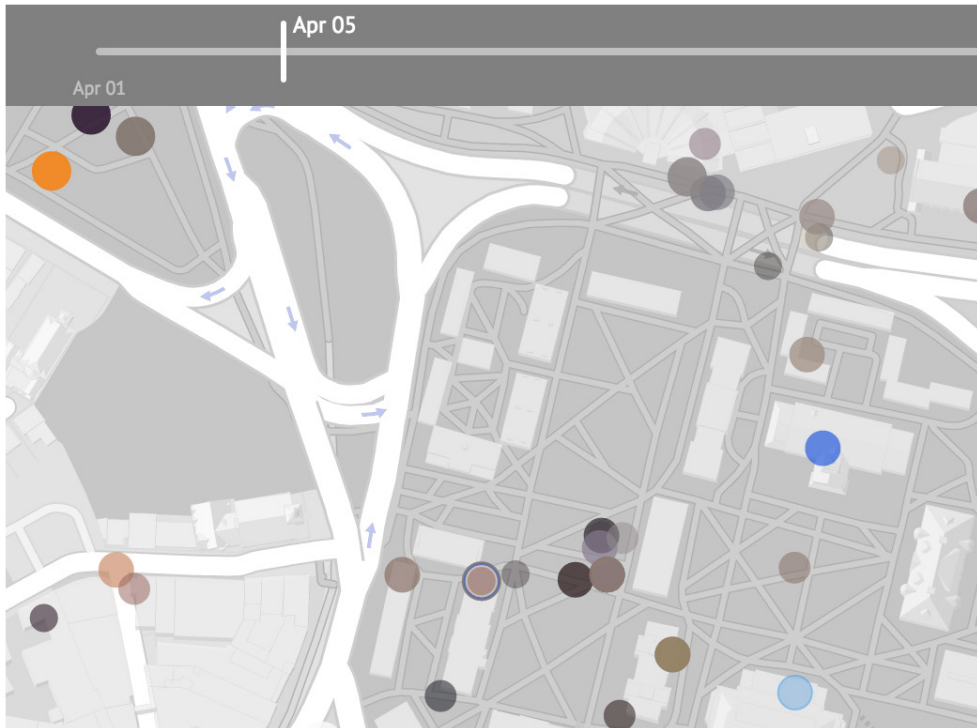
FORMING THE LAYOUT

Applying secondary graph canvases and organizing the page layout



USING REAL DATE

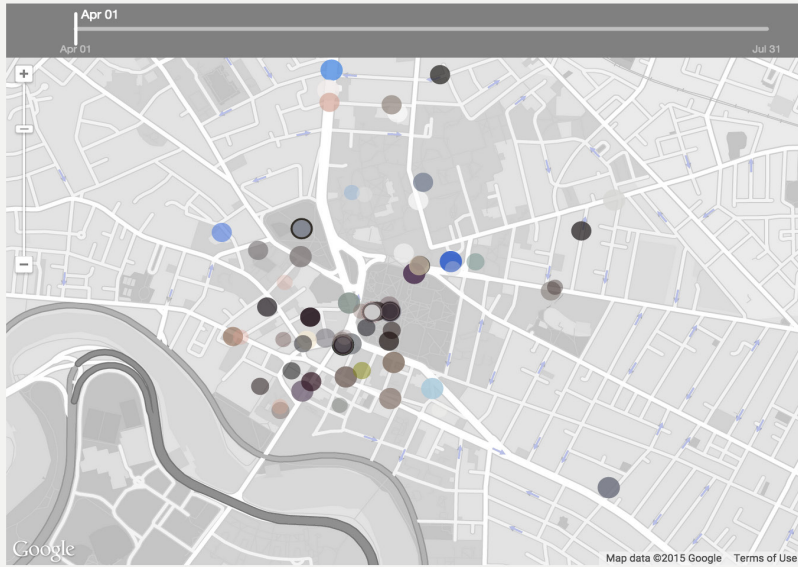
Mapping of all our thus far collected data (no date filtering applied)



DATA FILTERING

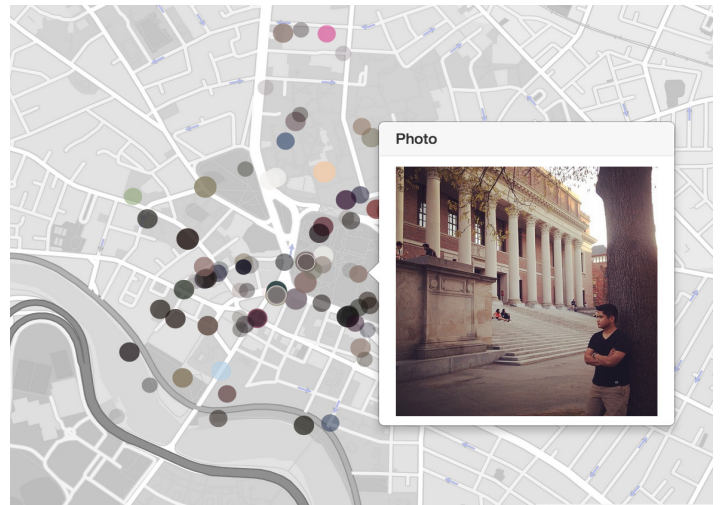
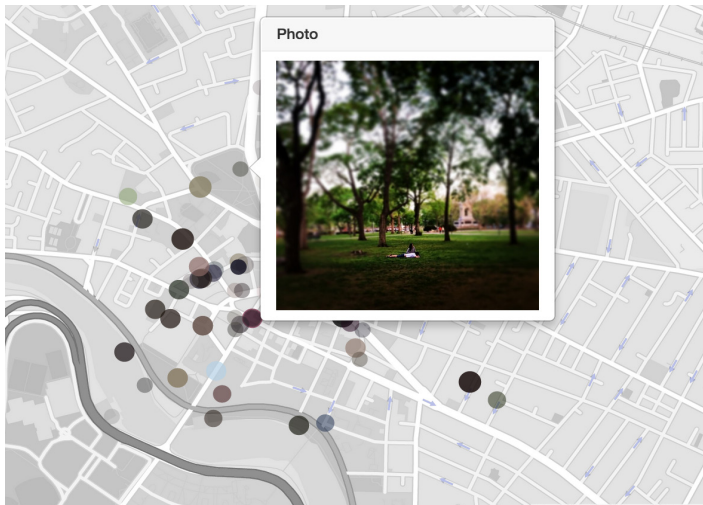
Application of slider that filters the data according to date. Reducing size & opacity of node as the difference from the slider date increases (1 week window)

Landscape Through Color: a Harvard Campus Case Study



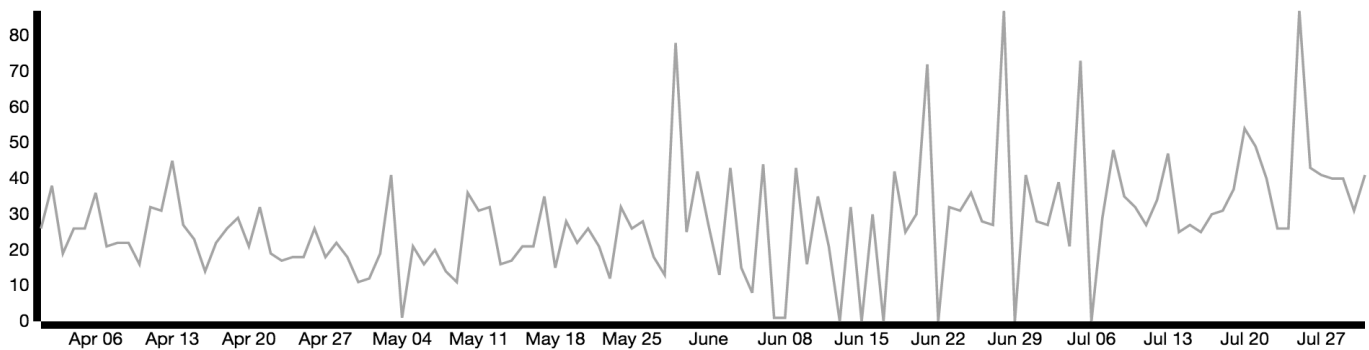
LAYOUT FORMATTING TESTS

Fonts, Background Color, Layout Placement



EXTRA FEATURES DEVELOPMENT

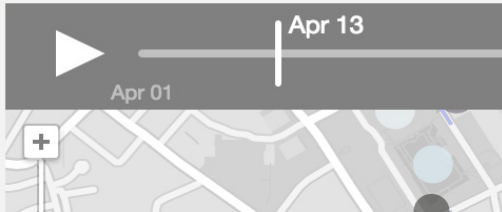
Implementation of Picture Pop-Up when node is clicked



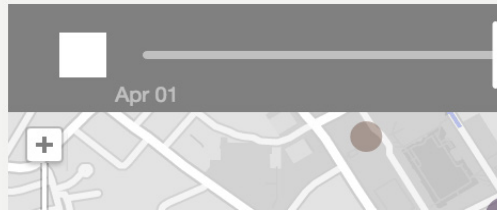
GRAPH DEVELOPMENT

Count of Pictures Per Day Graph

Landscape Thro



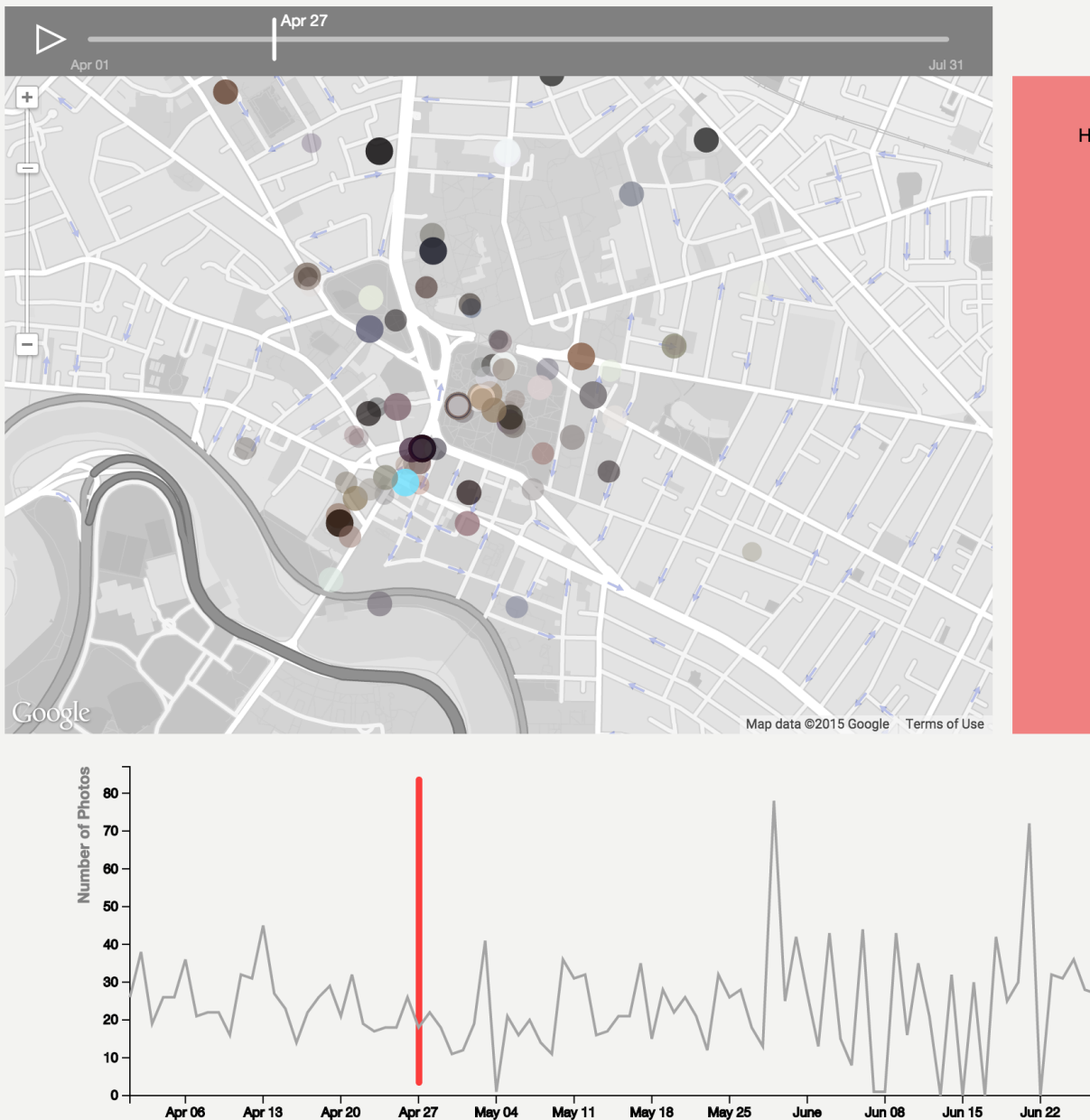
Landscape Thro



EXTRA FEATURES DEVELOPMENT

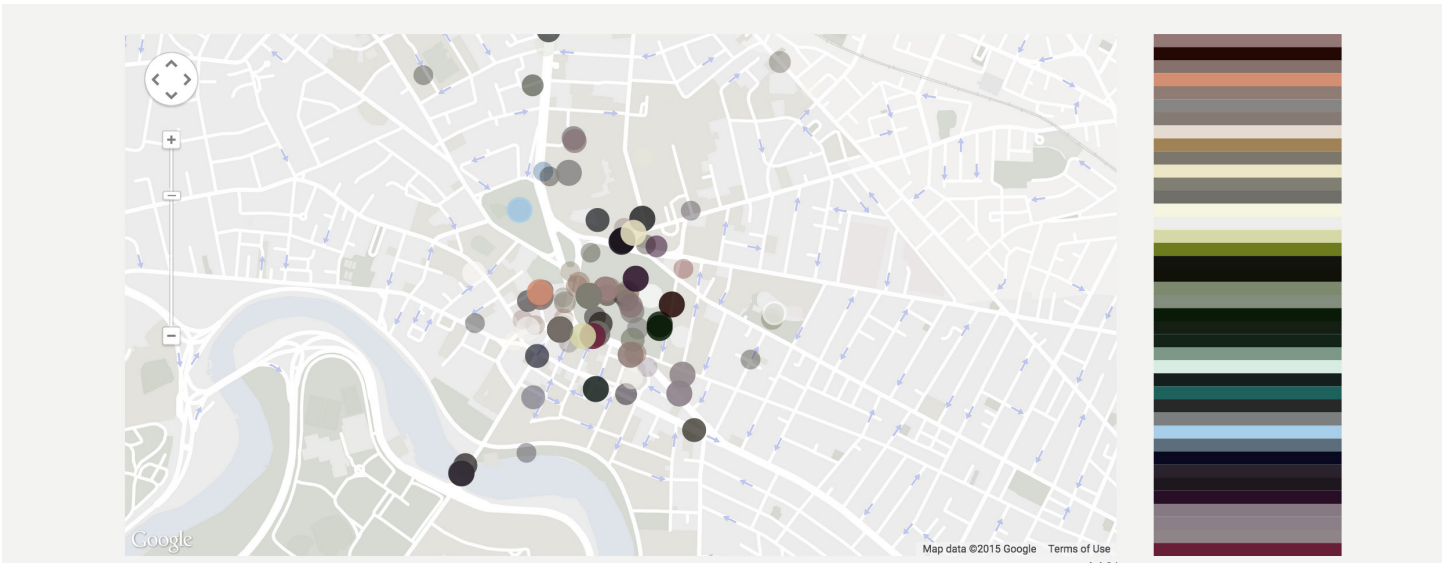
Adding Play and Stop Button in Slider Visualization

Landscape Through Color: a Harvard Campus Case Study



EXTRA FEATURES DEVELOPMENT

Adding an Indicator on the Count Graph that corresponds to the slider

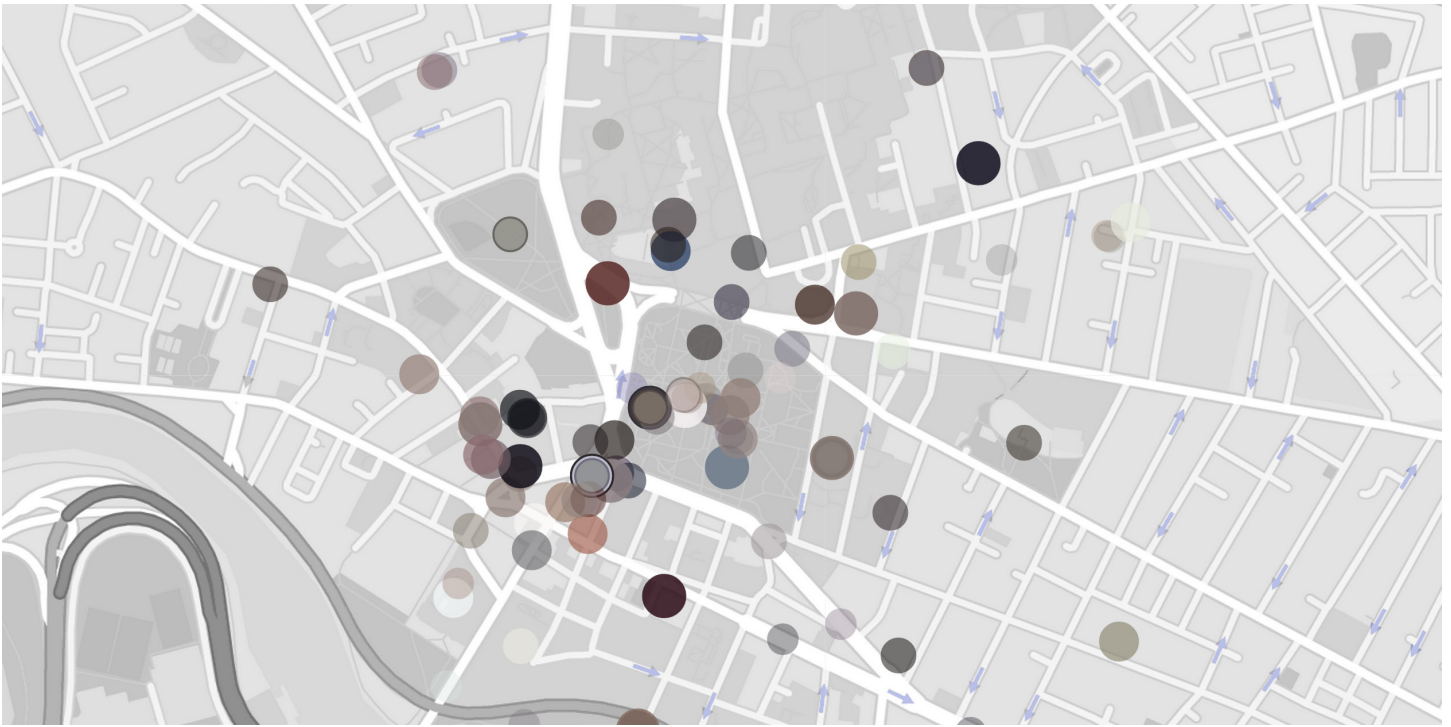


FIRST COLOR VISUALIZATION TESTS

Graph that shows the 1st primary color of each picture that was taken on the day the slider indicates

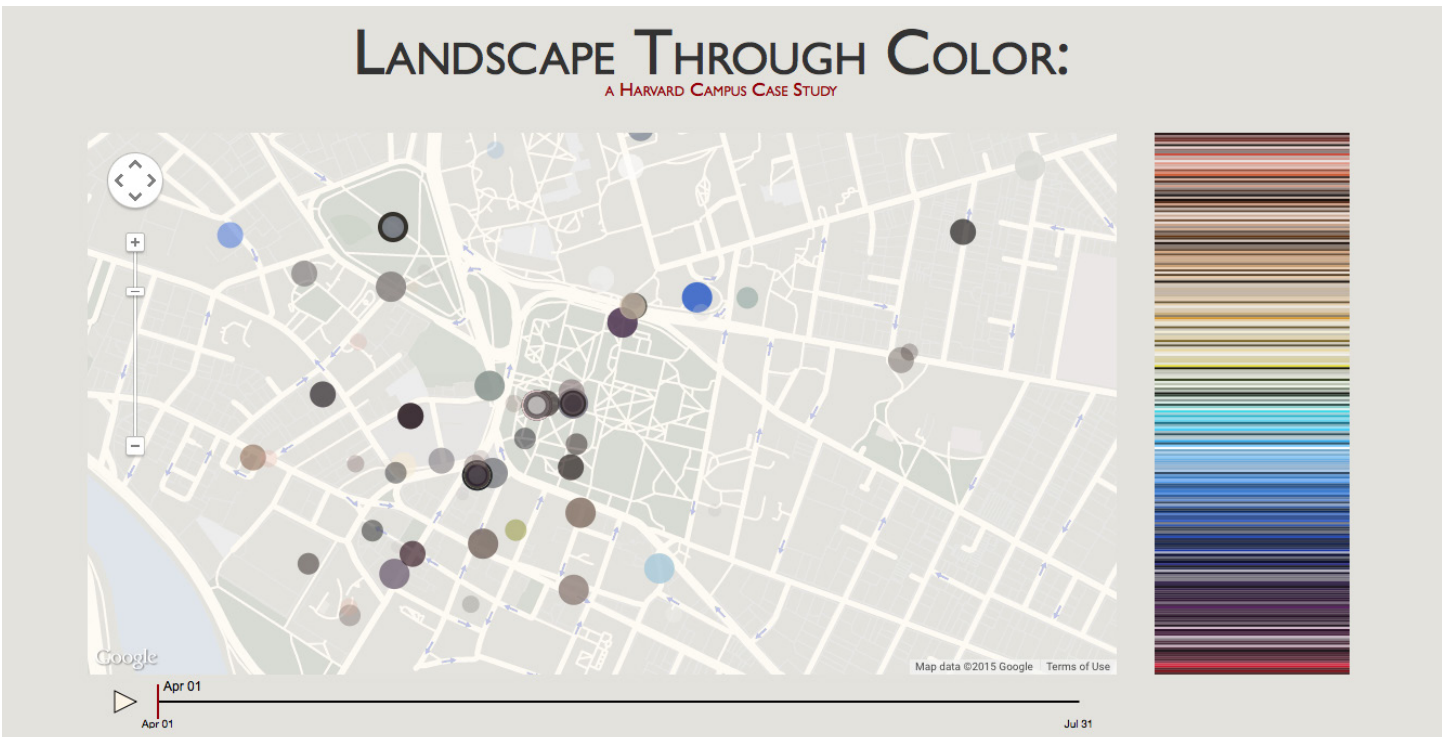
LAYOUT FORMATING

Adaptation of map colors, background, fonts and graph placement



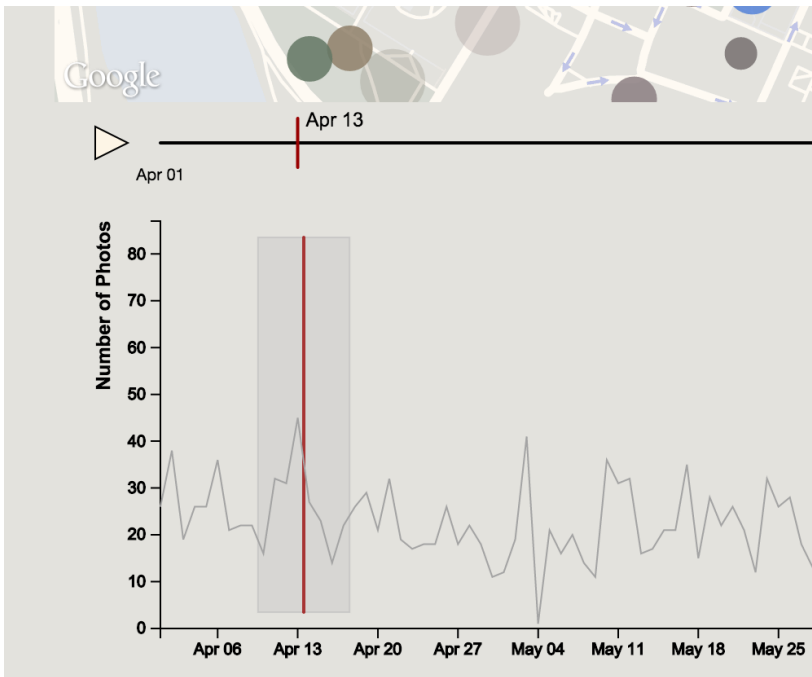
RETHINKING THE NODE VISUALIZATION

Making Nodes increase in size constantly (instead of increase & decrease as was primarily visualized)
 adaptation of date window, opacity and node size



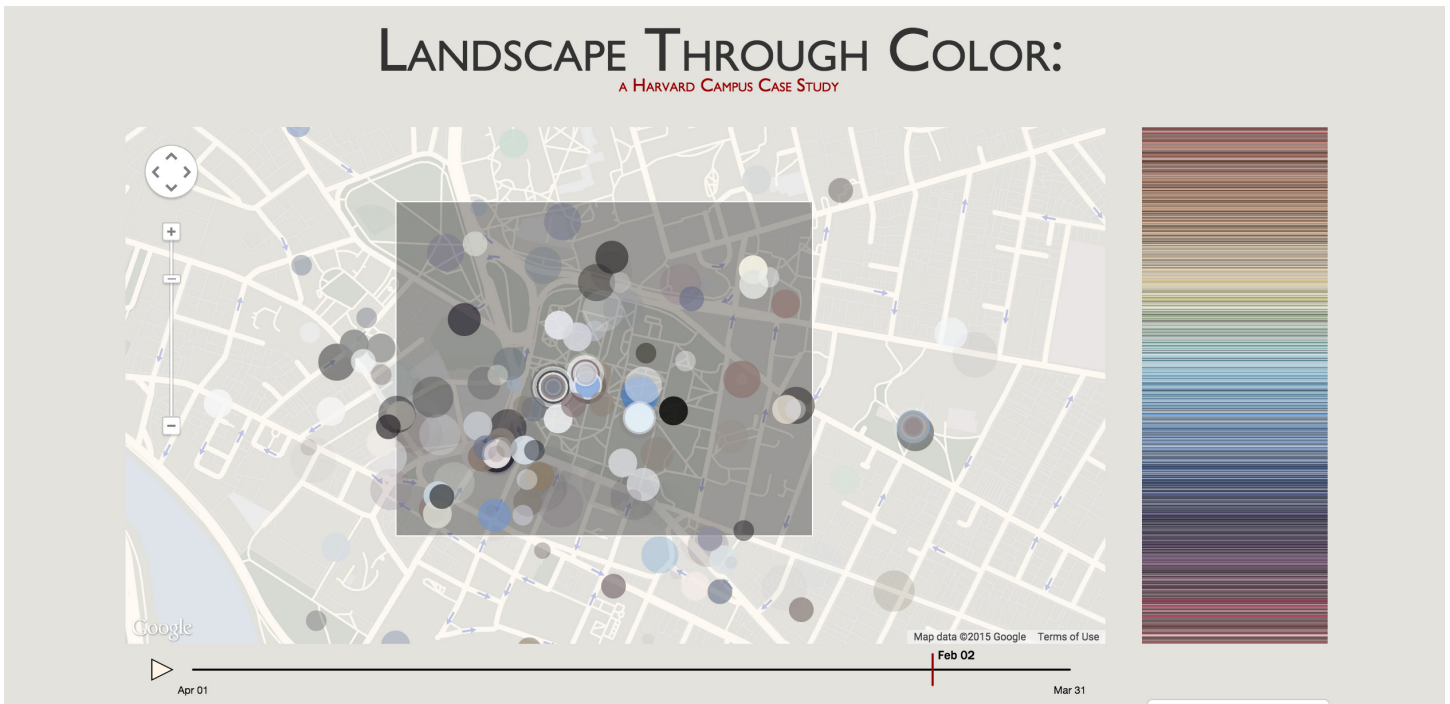
COLOR VISUALIZATION UPDATE

Visualizing all 10 primary colors of each picture in one graph



EXTRA FEATURES DEVELOPMENT

A rectangle of increased opacity shows the window of pictures to be visualized

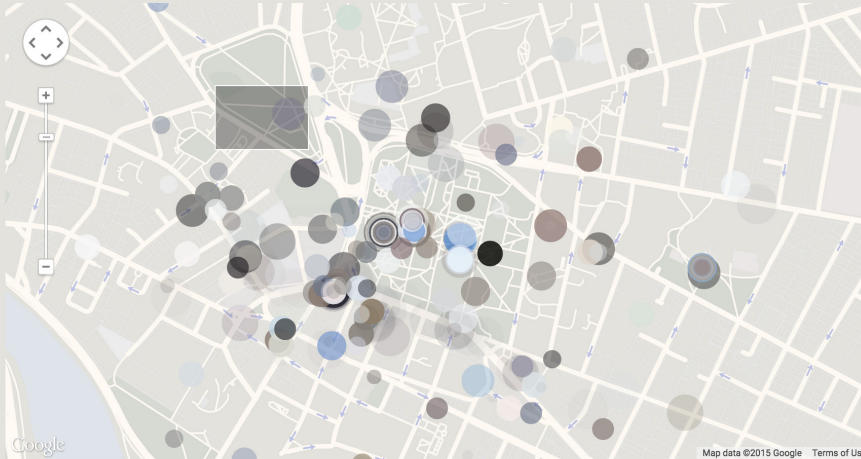


2D BRUSHING

2D Brushing on map visualization affects the colors of the graph (see next page)

LANDSCAPE THROUGH COLOR:

A HARVARD CAMPUS CASE STUDY



Apr 01

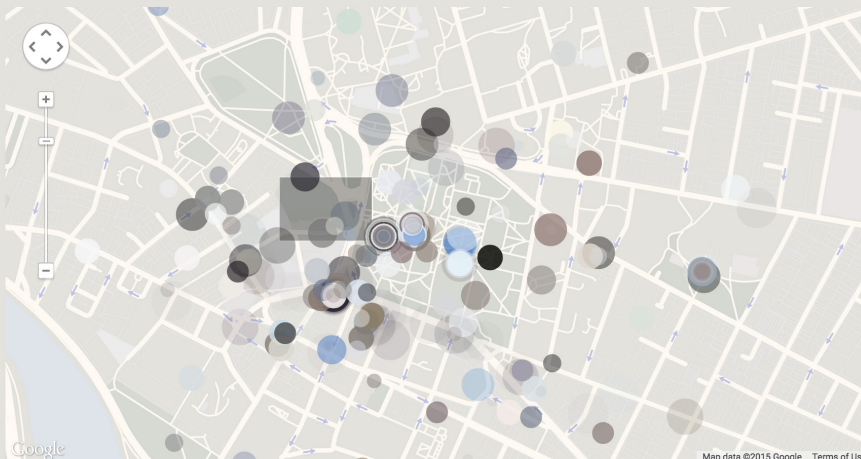
Map data ©2015 Google Terms of Use

Feb 02

Mar 31

LANDSCAPE THROUGH COLOR:

A HARVARD CAMPUS CASE STUDY



Apr 01

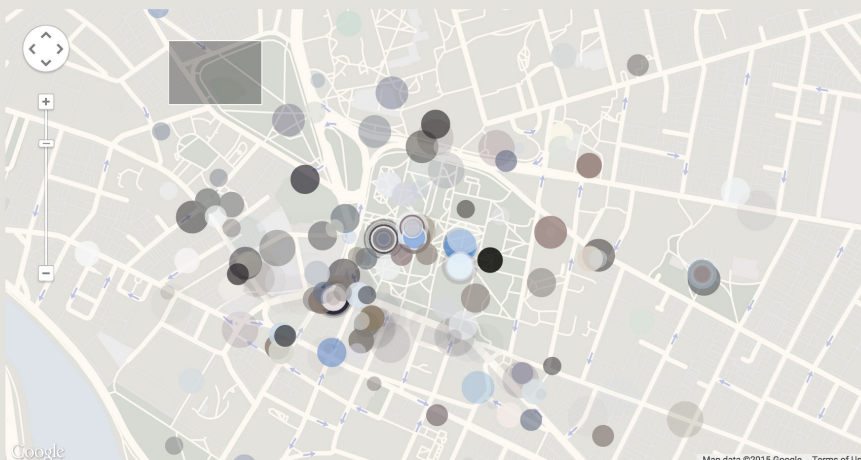
Map data ©2015 Google Terms of Use

Feb 02

Mar 31

LANDSCAPE THROUGH COLOR:

A HARVARD CAMPUS CASE STUDY

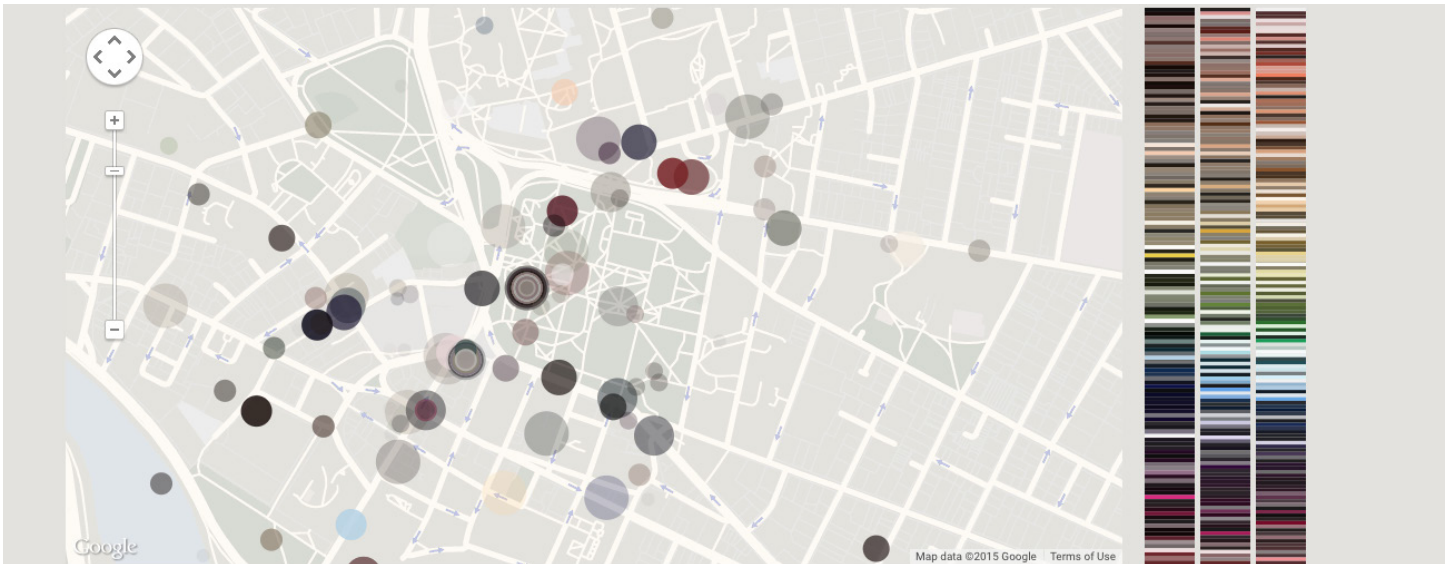


Apr 01

Map data ©2015 Google Terms of Use

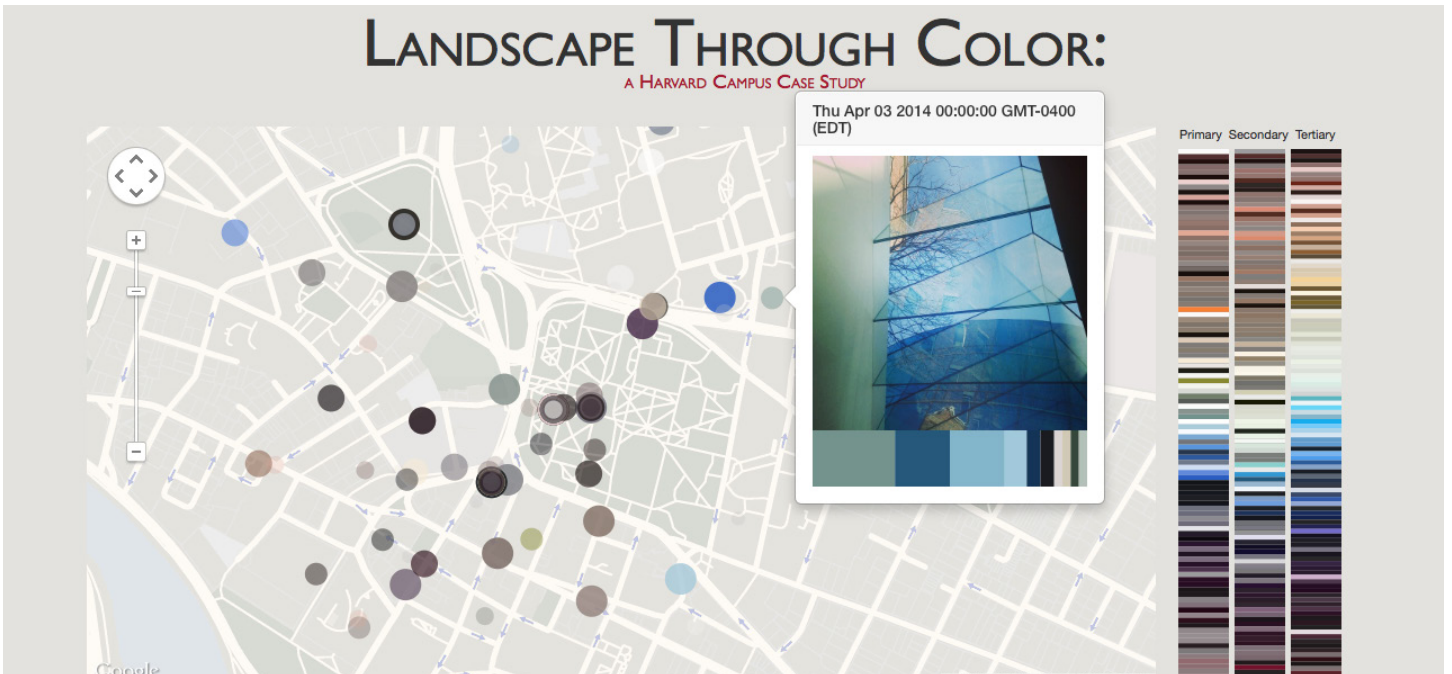
Feb 02

Mar 31



EVOLVING COLOR VISUALIZATION

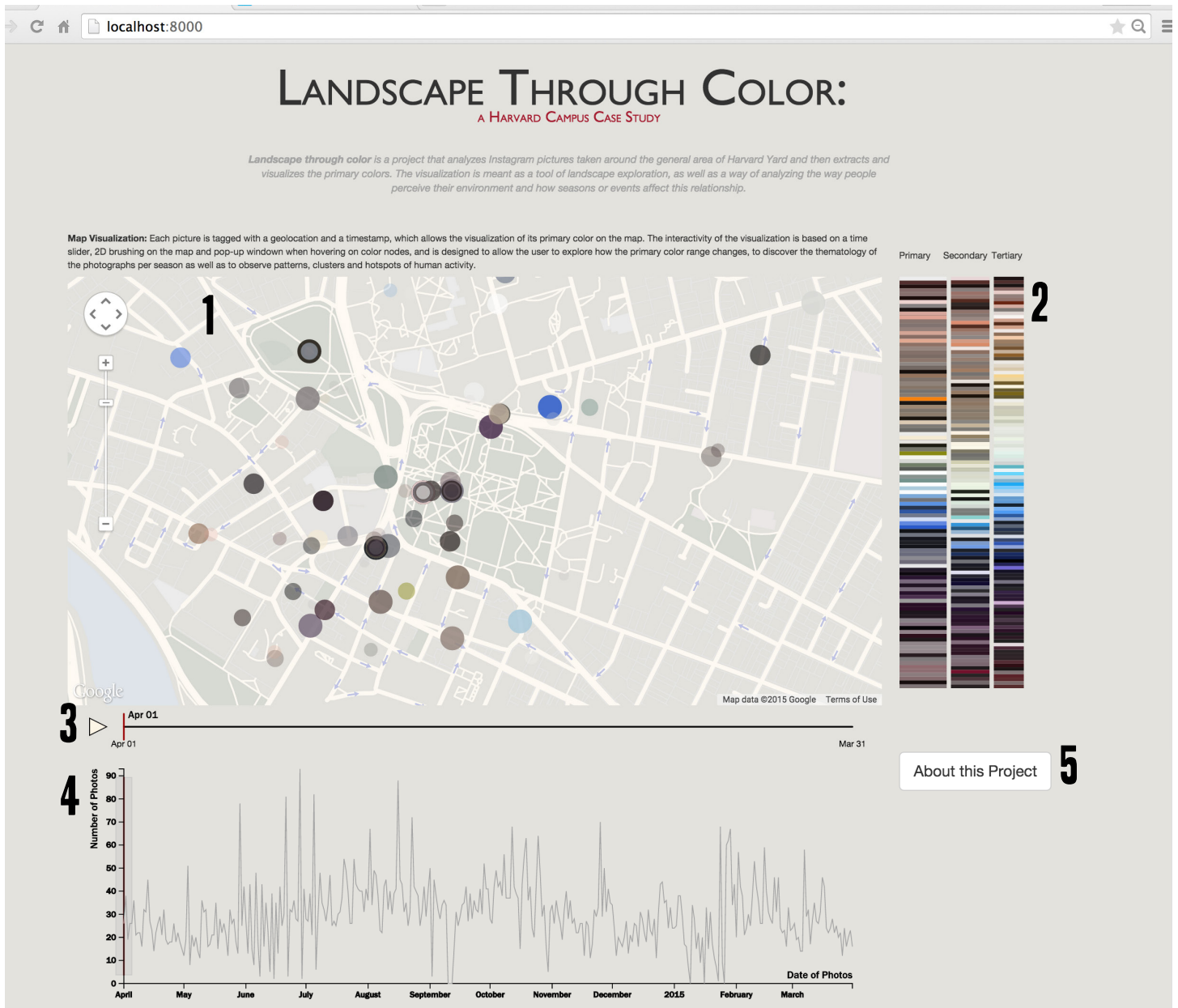
Creating Separate graphs for 1st, 2nd and 3rd primary color



EVOLVING COLOR VISUALIZATION

Adding a 10 color analysis under each picture pop up

FINAL IMPLEMENTATION



FEATURES

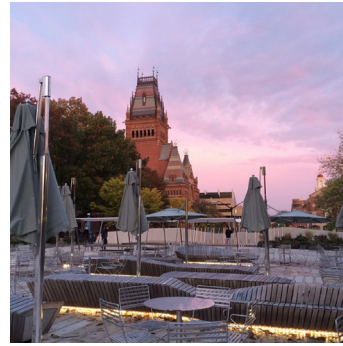
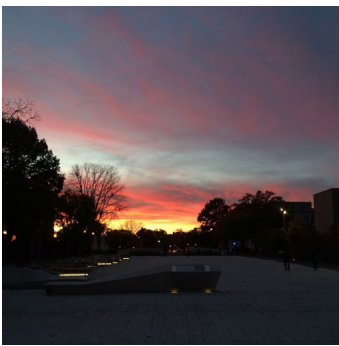
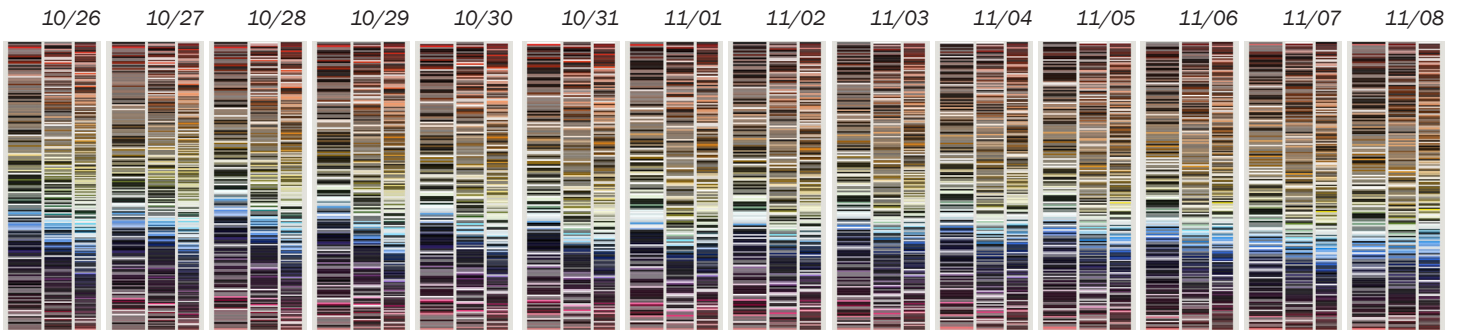
- (1) Geolocation mapping of color nodes
- (2) Graph of 3 primary colors of visualized pictures, sorted by hue
- (3) Time Slider
- (4) Count per day Graph
- (5) Video & Progress Book

INTERACTIONS

- (1) Hovering over color nodes reveals original picture with 10 primary color analysis
- (2) Hovering over color visualization graphs reveals original picture with 10 color analysis
- (3) Clicking on Time Slider or Count Graph sets the date accordingly
- (4) 2D Brushing on Map interactively alters color visualization

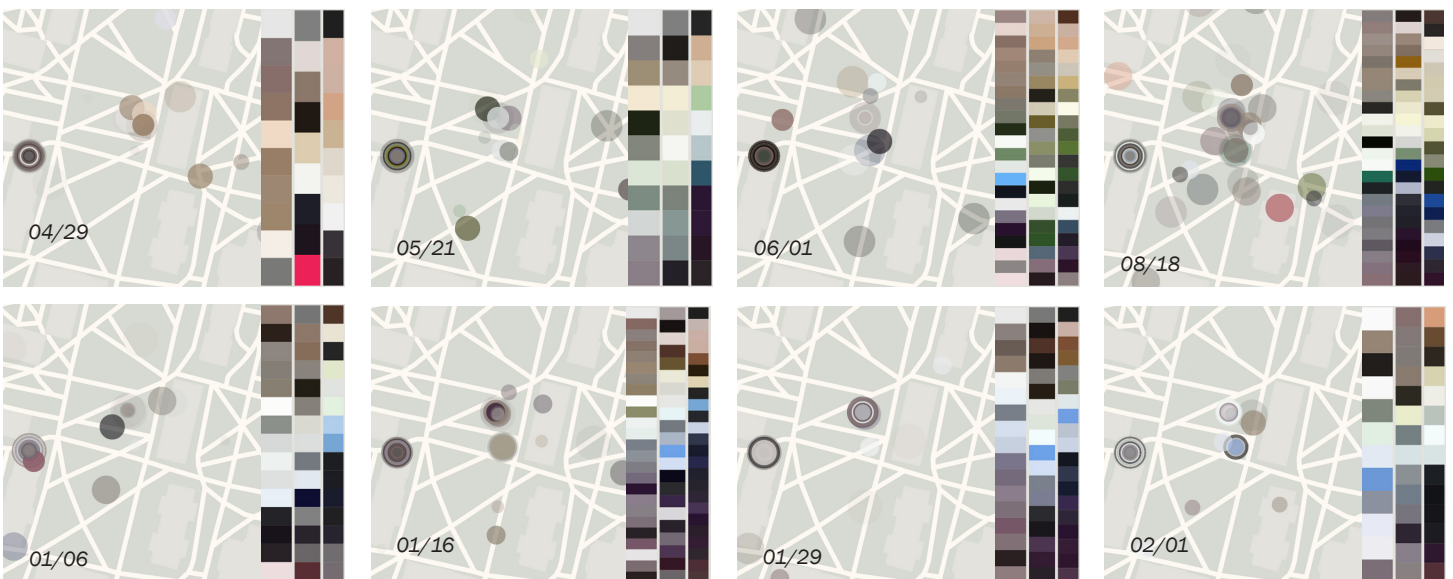
POINTS OF INTEREST

Examples of color analysis points of interest, color comparison, revealing patterns or discovering events.



COLOR HUE HINTS

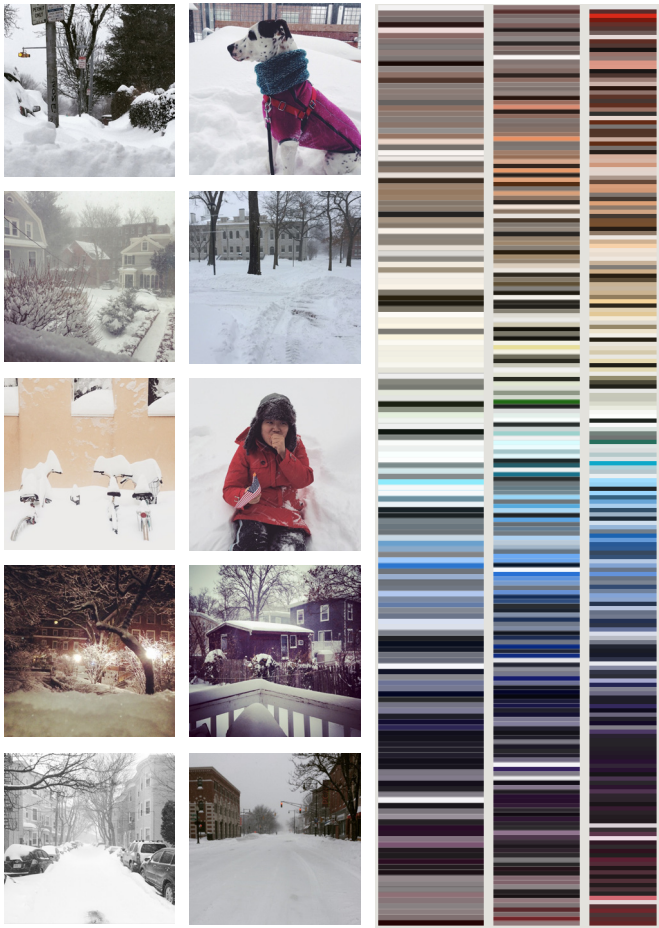
During the first days of November, an increase in pink and purple color hues was apparent, because of a beautiful sundown sky that was widely documented by people. Such color hue differentiations can give us clues on events or behaviors that cannot be otherwise predicted or traced.



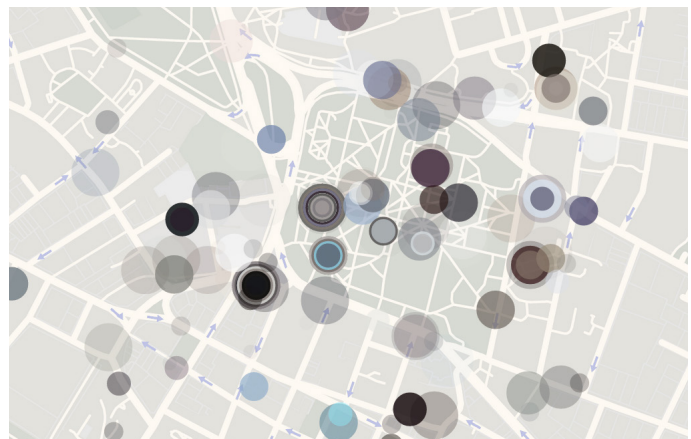
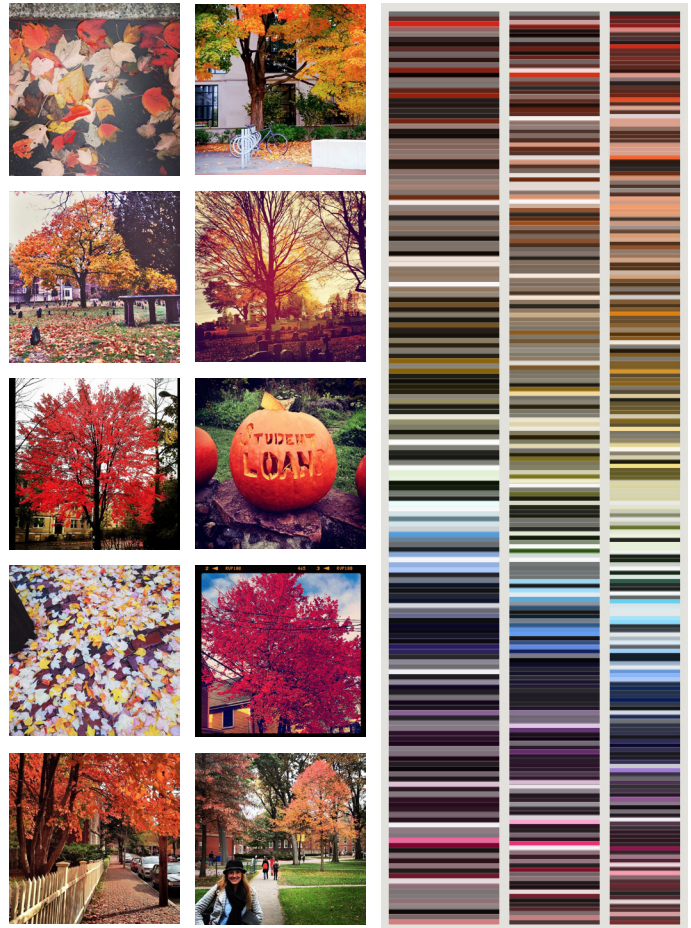
PERMANENT CLUSTERS

Some locations are particularly interesting because of the consistency of their popularity. These areas could be case studies of color change of a single theme, through the eyes of different visitors. (#JohnHarvardStatue)

january 2015



october 2014

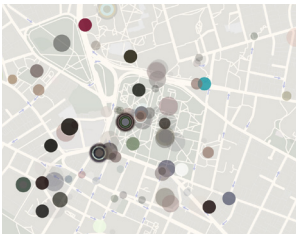


SEASONAL COLOR RANGE

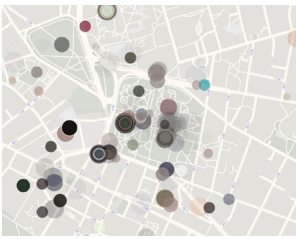
The most profound differences in color change occur when the natural elements take over and transform the landscape significantly. This also changes the behavior of people towards their environment, as we can clearly see from the alterations in photograph thematology.

In our examples, hues of red and yellow are prominent during the Fall, and trees become the real protagonists of the season to which Harvard Yard heavily bears witness. During Juno this past January, people were photographing the snow covered landscapes filling the color graphs with white hues, including pictures of ordinary objects and locations that were usually uninteresting to them, which we can clearly observe in the great diversion of the January color mapping.

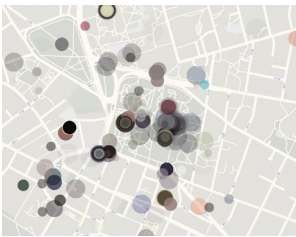
05/27



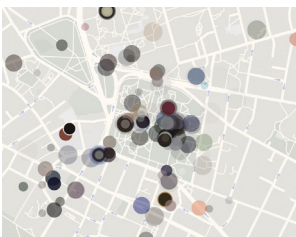
05/28



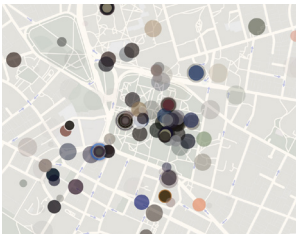
05/29



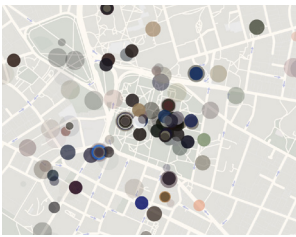
05/30



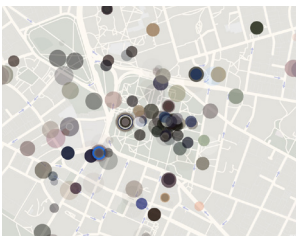
05/31



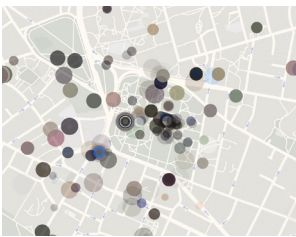
06/01



06/02



06/03

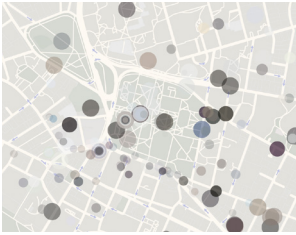


EVENT CLUSTERINGS

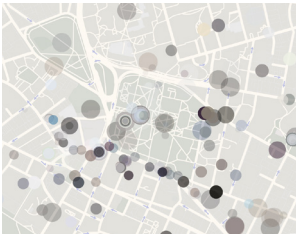
At the end of May Graduation day created some significant clustering of dark nodes in Harvard Yard, where the event was taking place. A location that is usually active, the area before Widener Library significantly increases its popularity during those days. (#CongratulationsClass2014)



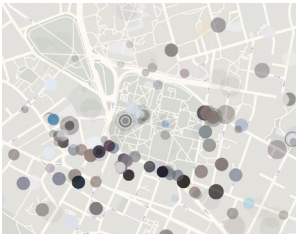
01/26



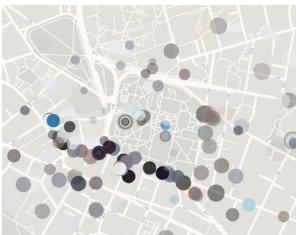
01/27



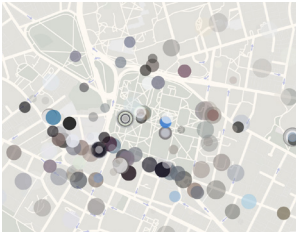
01/28



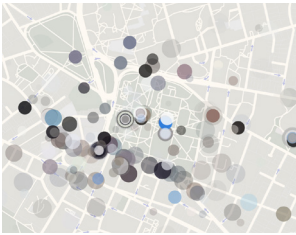
01/29



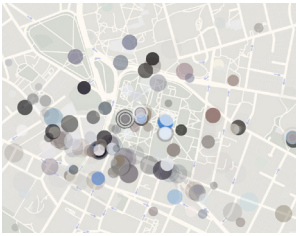
01/30



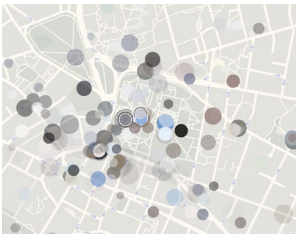
01/31



02/01

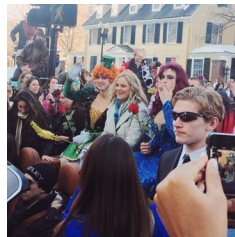
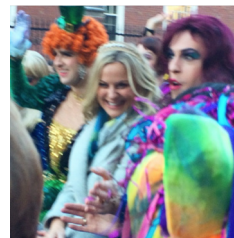


02/03



AMY POEHLER VISITS HARVARD

On January 29th, the Massachusetts-born comedian Amy Poehler parades through Mass Av to celebrate her announcement as Hasty Pudding's 2015 Woman of the Year. The enthusiasm of people is clearly documented through the photographs and a clear path of Poehler's movements can be traced through the Instagram pictures. #AmyPoehler



CONCLUSIONS

Overall, the project of Landscape through Color has exceeded our expectations in terms of the visualization outcome, as well as the efficiency of the application in detecting patterns, peaks and clusters that help research human behaviour in terms of landscape, as well as the landscape itself.

Through the graphs and the time slider, the visualization offers a tool to discover points of interest to photographers, whether they are typical sightseeing locations or small seemingly unnoticeable objects. It helps discover trails or hotspots of pictures, and recognize the color mood, not only of seasons but using the brush element of locations as well.

Finally, the visualization offers a more detailed scope of photograph themes by showing the individual photos, allowing people to discover what draws our attention and more importantly what we choose to document from our environment.

As the project ends, we still think there is some more exploring as far as optimizing the visualization capabilities, since we have produced a number of ideas that were not realizable within the course's time schedule. In conclusion, we would like to express our hope that we will be given the chance to further develop this tool so as to provide a background for anyone who is interested in studying the implications of color documentation of our environment.

Thank you!

Alkistis Mavroei, Jonathan Taratuta-Titus, Steven Vasilakos