Instagram Post Data Analysis

Yanling He*
Computer Science & Engineering
University of Washington

Xin Yang†
Computer Science & Engineering
University of Washington

Xiaoyi Zhang‡
Computer Science & Engineering
University of Washington

Abstract

Because of the spread of the Internet, social platforms become big data pools. From there we can learn about the trends, culture and hot topics. This project focuses on analyzing the data from Instagram. It shows the relationship of Instagram filter data with location and number of likes to give users filter suggestion on achieving more likes based on their location. It also analyzes the popular hashtags in different locations to show visual culture differences between different cities.

Keywords: data visualization, instagram, filter, hashtag

1 Introduction

As "a picture is worth a thousand words", more and more people are sharing their daily life, personal interests, news and events using images on social platform. Instagram is such a platform which is a popular mobile photo sharing application. It launched in October 2010, and rapidly gained popularity, with over 100 million active users as of April 2012 and over 300 million as of December 2014. Because of these huge amount of information in the dataset of Instagram, and we are interested in analyzing and visualizing the patterns of them.

This project is focusing on analyzing Instagram data to learn about the culture differences between different places. This project analyzes how filter usage are distributed in 50 cities, which are the cities with most population in each state of United States. This give us the information about how filter preference and vision culture varies in different states. The project also analyzes the number of hashtags been labeled on the posts for each city. It shows the popular hashtags for each city which reveal the popular event or hot terms in different cities.

While you are using the Instagram, you may also have ever scrolled through the Instagram filter list back and force worrying about which one to use, and how to make more people like it. But since culture background and contents varies a lot from photo to photo, it is hard to make a simple suggestion that let everyone like it. Our project also analyzes the Instagram filter data based on location and like to help you solve this problem.

The goal of this project is to learn about visual culture and content differences to help catch both the artistic trends and event trends for different places. It can also help user to make better filter selection to improve photo quality and reach more likes.

---

*e-mail:heyl@cs.washington.edu
†e-mail:yx1992@cs.washington.edu
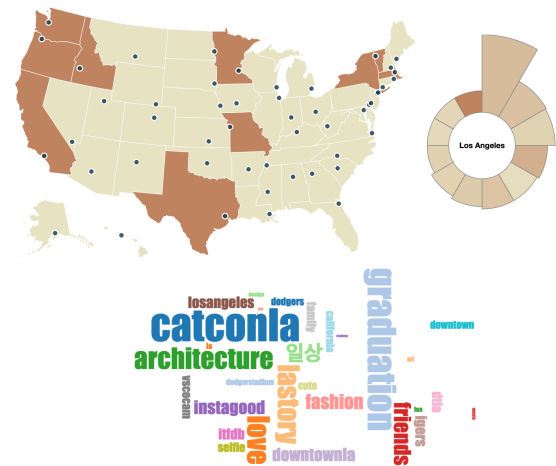‡e-mail:xiaoyiz@cs.washington.edu

Figure 1: *Visualization system design*

2 Related Work

Due to the popularity and the amount of data information, there are couple groups of people are also interested in the culture information reflected in the Instagram photo data. Hochman's Zooming into an Instagram city [1] [2] visualizes and analyzes samples from a data set of about 550,000 Instagram photos from New York City and Tokyo, by applying visualization and Cultural Analytics techniques. They show all the images in the collection. They download data based on latitude and longitude criteria and use specified tools to analyze the data. One interesting result of this paper is that there seem to be reoccurring spatio-temporal visual deviations in a specific time period and a set place. Based on those large sets of Instagram photos, they show how visual social media can be analyzed at multiple spatial and temporal scales. They also present analysis of social and cultural dynamics in specific places and particular times, and introduce new visualization techniques which can show tens of thousands of individual images sorted by their metadata or algorithmically extracted visual features. But as they only focuses on photo color and photo style, but our project cares more anout the label information and filter data. They shows the differences of visual styles in different time different cities, while we are analyzing the exact events and objects that people are captured and marked in different cities.

There are also some existing analytical tools for Instagram. Iconosquare (formerly Statigram) [3] provides useful statistics about Instagram. It can also respond to comments and monitor hashtags. Instagram-analytics [4] contains huge amount of raw instagram data for users. Instastats [5] is python scripts to pull data from Instagram API. Those work shows us what data information are available from Instagram and our project uses some of the tools to get our desired data for analysis.

3 Methods

We implement a software tool for user to navigate the Instagram data. The implementation details are divided into four parts:

Data preparation: getting data from Instagram and sorted our desired data into useable csv format.

Data analysis: we use table and chart to pre-analyze the data.

Function design: our system filters data based on location, and visualizes filter and tag data.

Visual style design: our system uses map to display location data, uses Aster chart to visualize filter data, and uses word cloud to visualize hashtag data.

3.1 Data Preparation

First, we get Instagram data from Instagram API which allows us to specify the scope of the access we are requesting from the user. It provides four types of scopes:

basic - to read data on a user's behalf, e.g. recent media, following lists (granted by default);

comments - to create or delete comments on a user's behalf;

relationships - to follow and unfollow accounts on a user's behalf;

likes - to like and unlike media on a user's behalf.

In our project, we are insisting in the basic and likes scopes including filter types, tags, likes, image link and location information. We get the data through API endpoint and use instastats python script to pull data. The example data is shown below:

```
"data": [{
    "filter": "Earlybird",
    "tags": ["expobar"],
    "likes": {
        "count": 35,
    },
    "link": "http://instagr.am/p/BUS3X/",
    "location": {
        "latitude": 37.780885099999999,
        "longitude": -122.3948632,
    }
}]
```

We found data collection process is painful. Instagram's API has unpredictable behavior. For example, even if we indicate the number of post we want in query, its API will still give us different number of result in each run. Thus, in order to get enough data, we need to run multiple times and remove duplicate results.

3.2 Data Analysis

In order to find interesting pattern from the raw data, we use python script to process the data. We first explore the relationship between filter and number of likes. We ignore the location and get the whole table of filter, number of photos with this filter, maximal likes, average likes and total number of likes. With the help of Tableau, we can see the following result in figure 11,3 and 4.

We can see that the normal filter appears dominantly. This is probably because normal filter is the default setting of Instagram. Therefore if people are not very familiar with Instagram, then they will hardly pick up advanced filters. So many filters use normal filter that most popular photos are using normal filters. If we remove the effect of normal filters, we can see that the curve of number of

Figure 2: *Likes and filters in Seattle*

Figure 3: *Number of photos in different filters*

photos on filters looks like an exponential function, which is an evidence that it may follow the power law...(line truncated)... In this way, we may design a recommendation system for filters: since normal filter is so common, we can help users to use other advanced filters, in this way users may obtain sense of accomplishment while achieving lots of likes.

We also study the relationship between the likes and the hashtag. We create the table of word, total like, average like and maximal like for each city. The following chart is the result for Seattle. We do not find interesting pattern in the chart. We think the main reason is that we do not have enough data. Most hashtag appear only 1 or 2 times, which is too few to analyze. However, we can still observe that hashtags with popular photos are meaningful, that is , we can see some kind of trend from the hot hashtags. This can be used to notice users what is happening now in the city.

3.3 Function Design

There are three main functionality, focusing on location, filter and tag data.
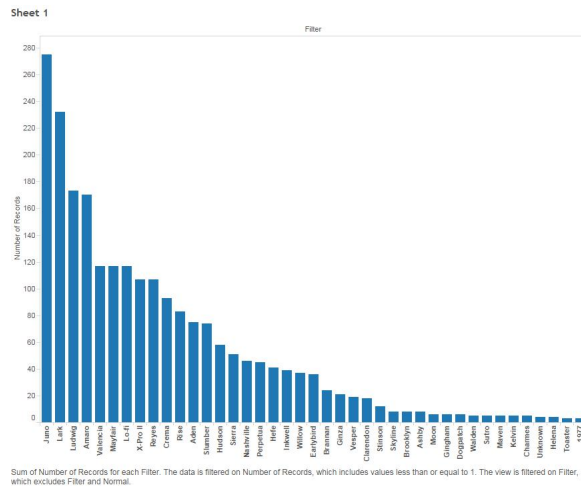
Sheet 1

Figure 4: *Number of photos in different filters. Normal excluded*
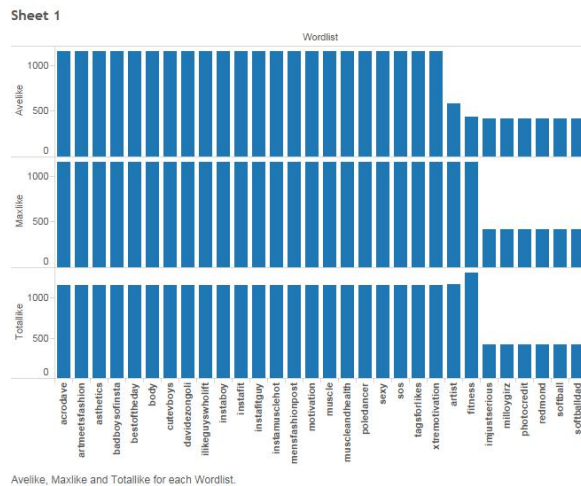

Sheet 1

Figure 5: *Hot hashtags in Seattle*

3.3.1 Location

User can click on dark blue dot, which is the city with largest population in a state. Then, the system shows the filter and tag visualization result based in the specific location. We want to use location as a filter for visualization results.

3.3.2 Filter

In the filter visualization result, we want to show the average likes of photos using these filters. We also want to show how many photos using those filters, which indicates the popularity of each filter.

3.3.3 Hashtags

In the tags visualization result, we want to show the average likes of photos using these tags.

3.4 Visual Design

We design our tool color theme using Instagram color patterns since we are analyzing Instagram data, this can provide users a more related and clear connection between the data information and the

tool. At the top of the page it shows a brief introduction of the project and overall functions to gives users a understanding about how to navigate and explore the tool.

3.4.1 Location - Map
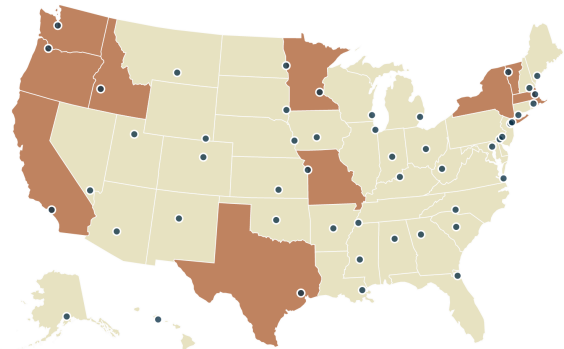
Figure 6: *Map design*

The project uses the map of United States. The color of the map is also chosen based on the Instagram color, the states with darker brown color are the top ten most Instagram posts states in the given time period, which is the recent past week. Lighter yellow color are the rest states. Since darker color presents more and higher density, so we choose darker color to represent more posts. Using dark blue color as the city dot is to form a contrast with the yellow and match the Instagram color theme.

3.4.2 Filter - Aster Chart

Figure 7: *Filter chart design*

When click on the city dot, an Aster chart[6] that shows the relationship between the filters and likes for the selected city shows up. Larger slice means more posts for that filter and smaller slice means less posts for that filter, because the size can be related to the amount information. As for the color, lighter color of the chart slice means less likes for that filter, and darker color means more likes for that filter. Most liked one uses the brown color and the least liked one uses the yellow color where the colors are match the color of the map. The color for the rest filters are the color within this range and equally scaled. When mouse hovering on the slice, the slice

turns into red color to use contrast color to indicate one filter type is selected, and a window pop up near by the mouse showing the detail information with the number of posts and likes and sample images for selected filter type. The color of the text information are using couple bright colors selected from Instagram logo which can gives the sense of users that they are viewing Instagram posts and draw some attention on the text because of the brightness.
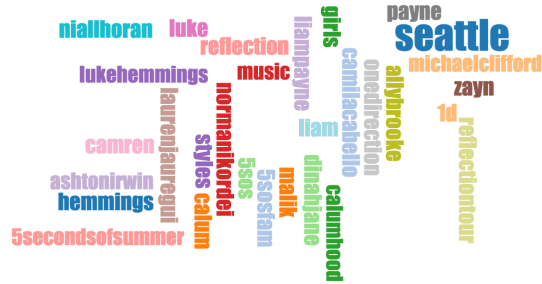
3.4.3 Hashtags - Word Cloud

Figure 8: *Word cloud design*

After clicking the city dot, there is also a words cloud[7] with tags showing underneath the map. Larger size of text means more posts with that tag and smaller size of text means less posts with that tag. The color of the text are all bright colors with large contrast to distinguish words from words and also for user easy to read.

4 Results

Our final work can be found in http://xiaoyizhang.me/512 . Basically, the top 10 states with most instagram posts can be seen directly as the darker parts on the map. Users can choose one city to see the hot hashtags people are using. Also users can see the amount and the popularity of different filters.

For the choice of filters, as we have mentioned, normal filter is most frequently used. In every city, the number of photos in normal filter is time times of any other filters. However, with respect to average, normal filter does not win everywhere. For example, in Charleston, the Lark filter achieves more average likes. So it is possible for advanced users to obtain better effect with interesting filters.
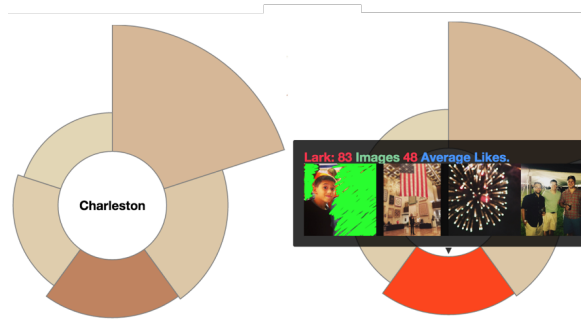

Figure 9: *Illustration for Charleston*

For the hot words, we can observe culture difference between different cities, Of course, the name of the location is a quite popular hashtag; however, if you check the word cloud of LA, you can find "graduation" and "fashion", which implies LA is a energetic city; on the other hand, for New York, you can see the names of a lot of Japanese animations. One possible explanation is that New York may be holding an anime exposition at that time..
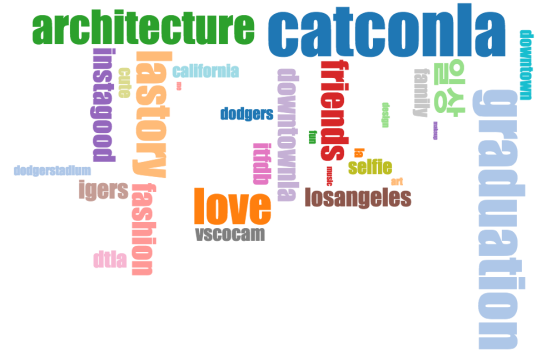

Figure 10: *Illustration for LA*

Figure 11: *Illustration for NY*

5 Discussion

5.1 Results Discussion

We present our work in the poster session. We are happy to see that all of our audiences find great pleasure in our work. Out of our expectation, they seem to be very interested in the word cloud. It is amazing to see what is happening in different cities, and it is quite interesting to see the different culture directly. They are also interested in the recommendation for advanced filters. Some people say they may apply the recommendation to become fancy.

However, they also express some confusion on the design. Most people asked about the meaning of different colors on the map. This is reasonable, since the color is the most notable element in the design. They also asked about the meaning of the pie chart of the filters. We think this is because we do not provide enough text explanations. We have modified our final design to meet the requirement of the audience.

5.2 Future Work

Since not all the photos are labeled with hashtags and not all the hashtags are correctly showing the content in each photo, using computer vision to analysis the real photo content, the style of the scenes and the major color theme may have stronger correlation with the filter types.

As the time changes people's visual preference may also changes, so the preference of filters may shifts as the time changes, we can learn the relationship with filters, likes and time to learn how visual preference changes and give out more current filter suggestion.

Since all the location analysis are based on the United States, so culture variety may be less between each cities, to extend the data to world based to learn some culture different between continents may give us more meaningful data. But world-wise spread of the Instagram usage may be the limit of this extension.

6 Conclusion

We have completed a visualization project on Instagram data. We study the relationship between the likes and the hashtags, location and filter, which many people are caring about. We use Tableau to visualize and analyze the raw data which has been processed to a easy-to use form and find interesting patterns. Then we use d3 to create a webpage interactive data visualization work. We combine multiple tools including datamap, aster plot and word cloud.

Compared with current existing Instagram analyzing tools, instead of focusing on individual user, we focus on big data on the whole Instagram community. We divide the data according to location, and in this way we detect culture different in different cities. This method can also be used in social science to study popular trend by collecting social network information. Also, we design a recommendation system, which is able to give general suggestions on choice of filters based on location. This may be improved with other advanced techniques like machine learning and computer vision techniques.

To summarize, it is amazing to visualize social network information data on real geographical map. This kind of work may be important in future development of data visualization and social science.

References

[1] Hochman, Nadav & Manovich, Lev. (2013) Zooming into an Instagram City: Reading the local through social media, *First Monday.*

[2] Hochman, Nadav & Schwartz, Raz. (2012) Visualizing instagram: Tracing cultural visual rhythms, *Proceedings of the Workshop on Social Media Visualization (SocMedVis) in conjunction with the Sixth International AAAI Conference on Weblogs and Social Media.* (ICWSM–12), (pp. 6–9).

[3] Key metrics about your Instagram account
http://iconosquare.com/

[4] Free Instagram User Report
http://simplymeasured.com/freebies/instagram-analytics

[5] Instastats
https://github.com/rldw/instastats

[6] Aster Plot in D3
http://bl.ocks.org/bbest/2de0e25d4840c68f2db1

[7] Word Cloud
https://www.jasondavies.com/wordcloud/