Instagram Post Data Analysis

Yanling He, Xian Yang, Xiaoyi Zhang

Department of Computer Science & Engineering, University of Washington

Introduction

Problem:
This project analyzing the relationship of Instagram filter data with location, number of likes and hashtag to give user filter suggestion to achieve more likes based on user's location and the photo content, and analyzing visual culture differences between the cities. It shows three types of data relationship: first, filter usage based on different cities; second, in each city the number of likes of each filter; third, in each city what hashtag are been labeled most.

Motivation:
You may 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. To solve this problem our project analyzes the Instagram filter data based on location, likes and hashtag.

This project analyzes how filter usage distributed in 50 cities, which are the cities with most population in each state of United States, and it also shows the number of likes for each filter. This shows the filter preference and how culture varies at different places. It also analyzes the number of hashtags been labeled on the posts for each city.

The goal of this project is to give useful filter suggestion and showing visual culture and content differences for different states.

Future Work

Computer Vision Analysis
- 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.

Relationship with Time
- As the time changes people's vision 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.

World Map
- 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.

Reference

D3.js: http://d3js.org/

Data Maps: http://datamaps.github.io/

Word Cloud: https://github.com/jasondavies/d3-cloud

Aster Plot: http://bl.ocks.org/bbest/2de0e25d4840c68f2db1

Data & Techniques

Data Preperation

Instagram API
- Instagram endpoint API returns location, filter and likes data for every post
- Averagely scripting 2,000 - 3,000 posts data based on the 50 different cities' location

Data Processing

ID	Created At	Filter	Number of Likes	HashTags
1002633384852124593_1236245955	2015-06-08 06:00:35	Normal	2	[Tag: gratitude, Tag: serenity, Tag: natureischurch]

- Feature selection: for each city, selecting time, filter, number of likes and hashtags as the features for each post.
- Pre-analysis: using the Tableau pre-analysis the relationship of filter location, likes and hashtag to a make sure there are some correlation between those data.
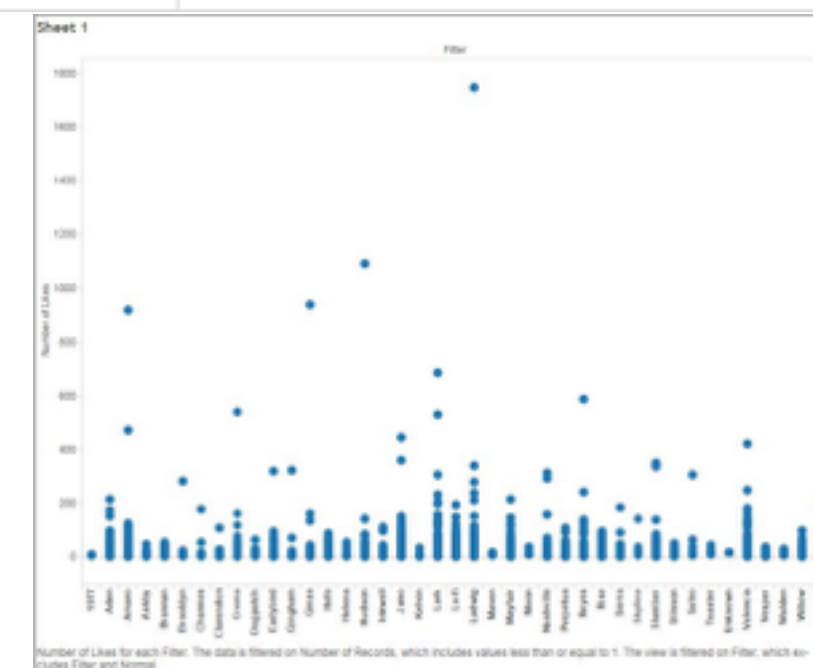
Techniques

D3.js
- JavaScript library for manipulating documents based on data

Data Maps
- Customizable SVG map visualizations for the web in a single Javascript file using D3.js
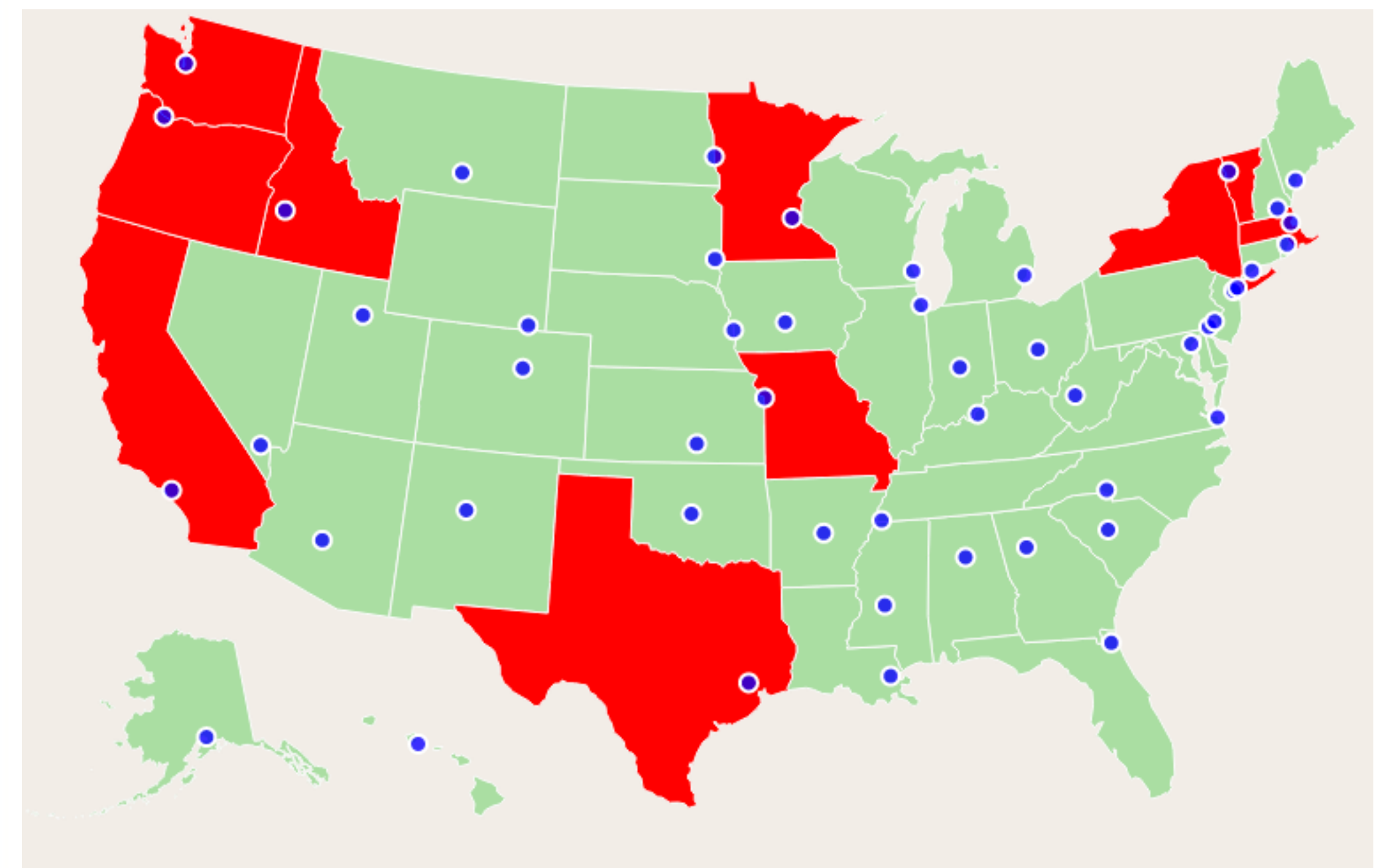
Word Cloud
- A Wordle-inspired word cloud layout written in JavaScript. It uses HTML5 canvas and sprite masks to achieve near-interactive speeds.

Aster Plot
- Aster plot displays pie slices using D3.js

## Approach Details & Results

Map Navigation

City Dots
- The city with most population of that state
- Mouse click the city dot, a PI chart shows up giving the relationship of filters and likes at that city and a word could shows up showing the number of some hashtags been labeled in that city

Color
- Red color states are the top ten states with the most Instagram posts
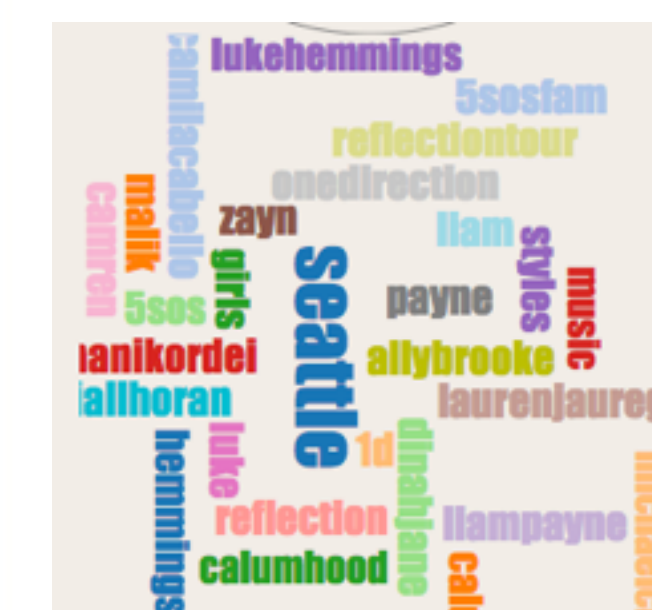
Pi Chart Design
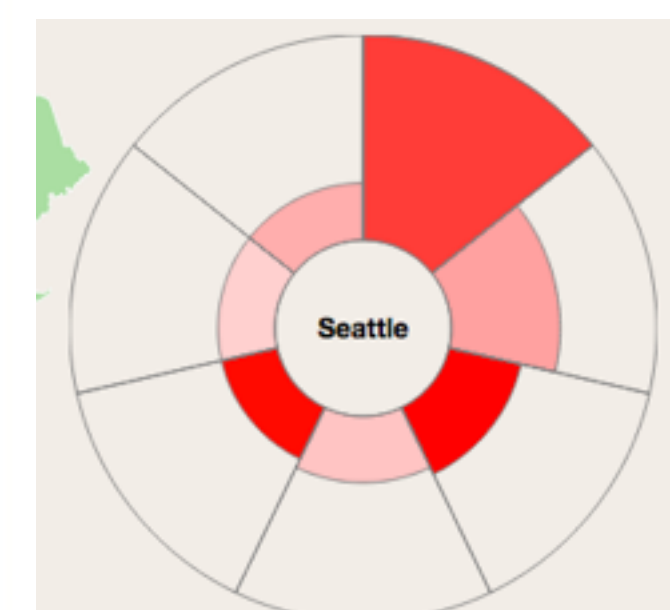
Pi Slice Size
- Larger size of slice means more posts for that given filter type; Smaller size of slice means less posts for that filter type for the selected city.

Color
- Darker color means more likes of that filter; lighter color means less likes of that filter for the posts of the selected city

Hashtag Cloud

Hashtags
- Hashtags that have been labeled on the posts for the selected city
- Bigger size means more posts are labeled with that tag; smaller size mean less posts are labeled with that tag in the given city