#InstaMAP

Chiew Jun Hao, Nguyen Bao Ngan, Zhang Hao, Zhang Zhexian

Introduction

The use of social media platforms, notably Instagram, has risen sharply in Singapore as the popularity of highly-visual channels continues to grow (Figure 1).
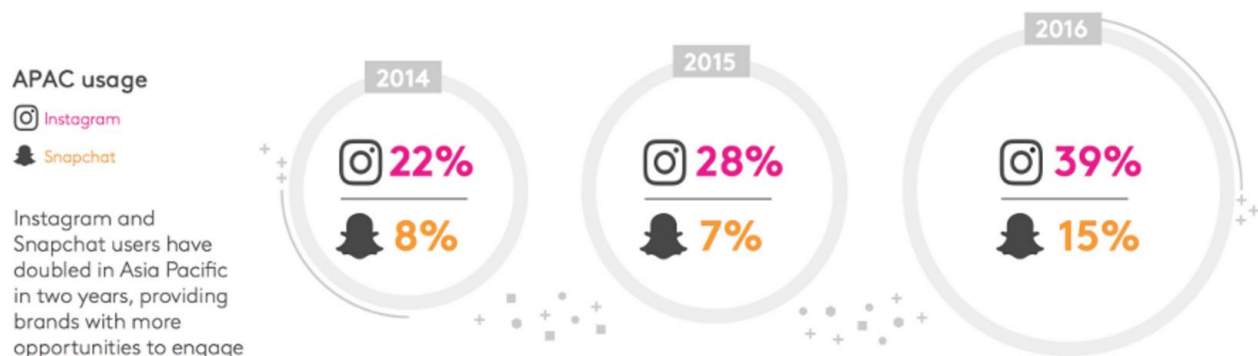

Figure 1. APAC usage of Instagram and Snapchat 2014-2016[1]

There are predictions on "Instagram-worthy Places in Singapore"[2] but no systematic analysis has been conducted to verify the claim.

Our project not only maps both the temporal and spatial shift of instagram users in Singapore since the introduction of this social media in 2010, but points out the clusters of most posted and most liked hot spots.

Potentially, our project benefits stakeholders from tourists to residents, from government to business owners, by providing insights and opportunities to target and engage the ever-growing Instagram audience (see Appendix).

Visualization

In our final map product, we aim to present an interactive web interface with both horizontal sliding and vertical zooming features.

---

[1] Infograph: http://connectedlife.tnsglobal.com
[2] Source: http://thesmartlocal.com/read/instagram-worthy-places

The horizontal slider allows user to change the time frame by years, from instagram introduction year 2010, to current year 2017.

The 3-level of zooming features allows users to gather various information depending on the level selected, as explained below (Figure 2):

Level 1

Overall number of posts per area and its changes over time. The user then selects the desired region to go for Level 2.

Posts density
Lowest   Highest

North
North East
West
East
Central

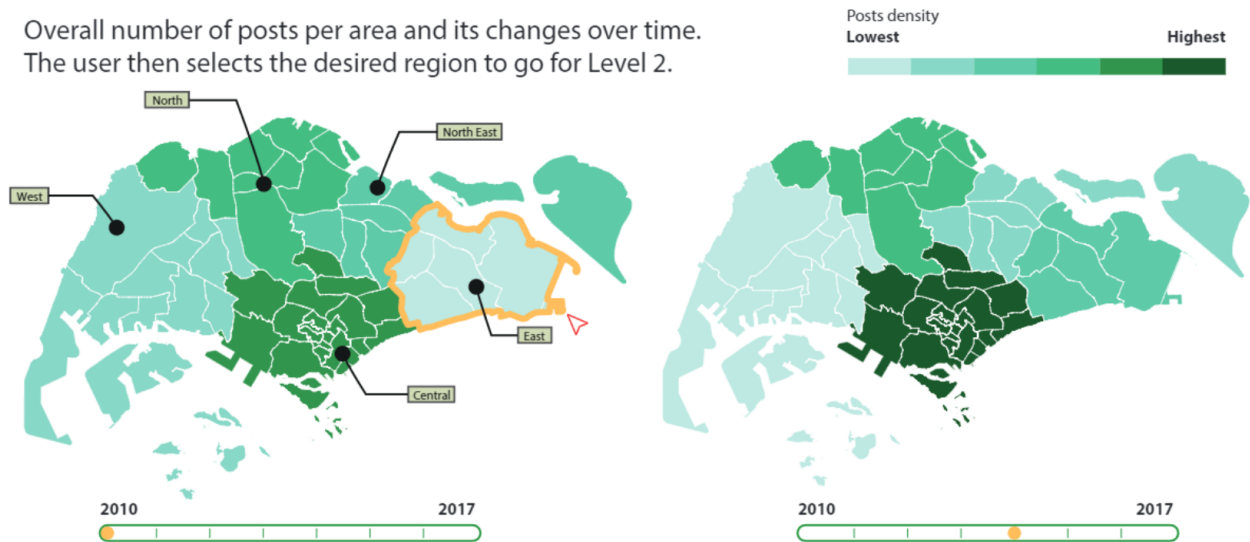2010   2017   2010   2017

Figure 2. Mock up of our final project (Level 1)

When one of the region in level 1 map is selected, the user enters the 2nd level (Figure 3):

Level 2

The distribution of posts over the selected region and the most favourite locations based on the number of likes. The user then picks the pre-selected location to move forward to Level 3.
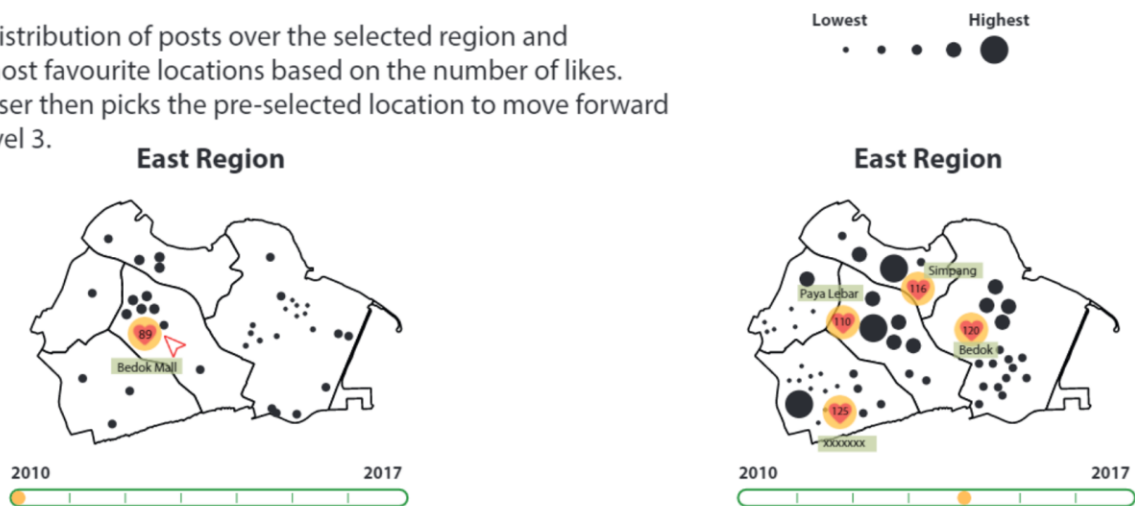

Figure 3. Zoomed in view of East Region (Level 2)

When a well-liked photo is selected, user enters level 3 (Figure 4):

Level 3

The top photos based on the number of likes within the favourite locations and the category statistics.
The photos are categorized into 3 groups, which are people, food, and scenery with its respective percentage.

Figure 4. Detailed view of Bedok Mall (Level 3)

Data Collection

How do we pull data from Instagram

Instagram publishes documentation on establishing connection and pulling data from their server. There are plenty of blogs and tutorial available online, and we will use R for data mining of Instagram post.[3]

To get the Instagram posts in Singapore, we need to do it in two main steps:
1. Obtain all the location IDs in Singapore
2. Retrieve media data in each location ID

If we want to get location ID from a specific region, we need to specify the latitude and longitude coordinates and a radius. The API will then return all the location IDs within the distance we specified.

The limitation of the API is that the maximum radius we can set is 750m, thus, we could estimate the number of API calls we need in this step by calculating the area we need to cover.

The area of Singapore is 719.1 km^2 and one API call covers pi*(0.75km)^2 = 1.76km^2, therefore we need to perform 719.1/1.76 = 408 API calls to get all the location IDs in Singapore. If on average n location IDs are returned in each call, we need to perform 408*n number of API calls to get the media data.

How do we optimize data pulling

[3] https://www.r-bloggers.com/analyze-instagram-with-r/
[4] stackoverflow.com/questions/7716460/fully-cover-a-rectangle-with-minimum-amount-of-fixed-radius-circles

The previous calculation didn't take account in the overlap of area in the first step. To minimize the overlapping area in each API calls, we can pick the coordinates in the following manner. The green area represents area we wanted to cover and each circle represents the area covered in each API call in the first step (Figure 5).
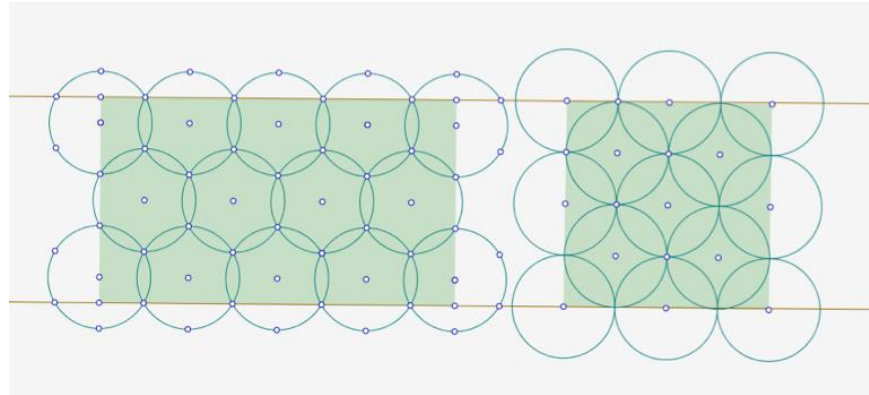

Figure 5. An example of minimizing area overlaps[4]

Data Analysis

The data that we will be analysing in a Instagram post:
1. Likes count
2. Created_time (date)
3. Location (lat, long)

Below are the list of the analysis that we will perform in this project:
1. To facilitate the transition of time in Level 1 visualisation we will categorise data by time period & normalize years of data less than 12 months (e.g. 2010 and 2017 data)
2. Calculate Instagram post density in each region:
   > Post density = number of posts / area of region

   This gives the map colour intensity based on the density in each region.
3. Rank the posts at each location based on the number of likes. Top 9 image will be displayed per location. To maximize the potential value of our calculations, we only conduct this analysis for locations with posts number above a certain threshold, i.e. only posts in popular locations are analyzed.

References

1. Instagram usage statistics: http://connectedlife.tnsglobal.com
2. Instagram-worthy Places in Singapore: http://thesmartlocal.com/read/instagram-worthy-places

3. R programming for Instagram data mining: https://www.r-bloggers.com/analyze-instagram-with-r/
4. Fully cover a rectangle with minimum amount of fixed radius circles http://stackoverflow.com/questions/7716460/fully-cover-a-rectangle-with-minimum-amount-of-fixed-radius-circles

Appendix: impact of our project by audience

Local: what's new in hometown, find new and popular photogenic spots

Tourist: travel like a local, find popular spots among locals

Singapore Tourism Board: good way to evaluate popularity of spots and guide future planning

Companies: more targeted advertisement at popular spots

Other Social Media Companies: compare and contrast with instagram hotspot for meaningful analysis (e.g. find niche area of each social media)

Map Study Students: relatable topic and enough complexity in data collection, analysis, and visualization

Investors: predict the future popularity of Instagram (a product of IPO company Facebook)

Government: measure social media impact and regulate it

Psychology and Sociology Researchers: source of behaviour analysis

Company: if photos are diverse or all similar -> different opportunity for new products

Photographer: avoid cliche shots

Artist: inspiration for photo traveling/representation of Singapore