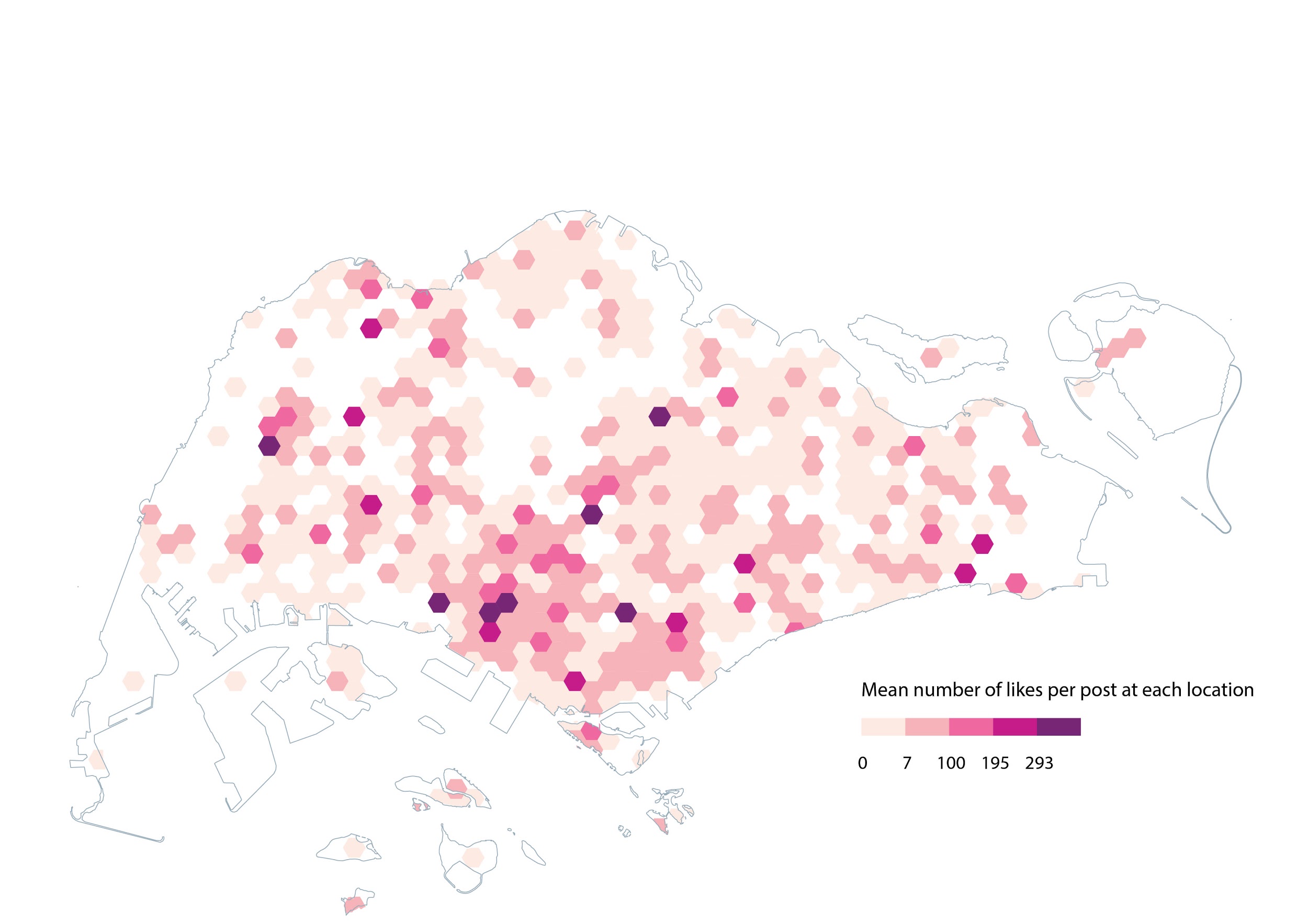
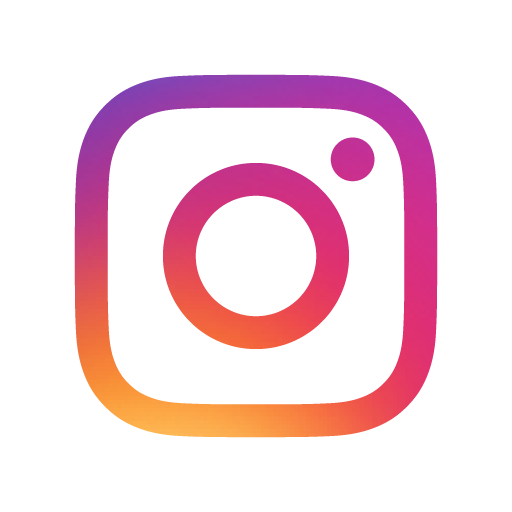


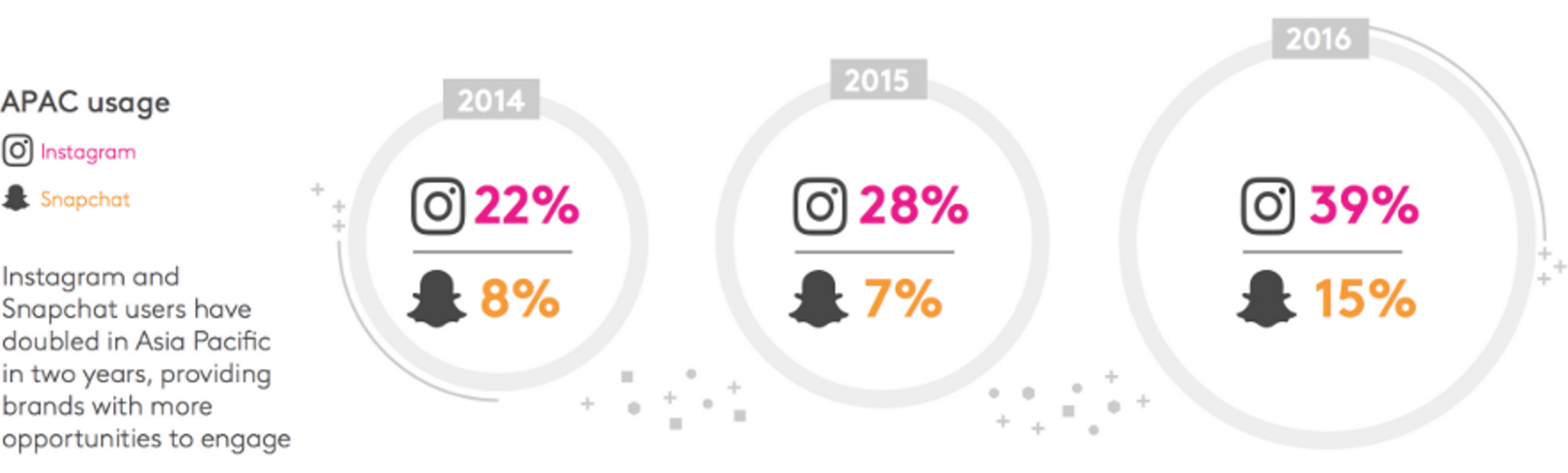
InstaMap

Team John Hall, 2017



# PROBLEM SCOPING

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]](#footnote-0)

There are predictions on “Instagram-worthy Places in Singapore”[[2]](#footnote-1) but no systematic analysis has been conducted to verify the claim.  
  
Our project aims to map the spatial distribution of Instagram posts across Singapore based on:

* The estimated post density in each location, and
* The estimated post favorability in each location.

With this information, we will be able to identify the clusters of Instagram hot spots in Singapore.   
  
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 A).

# DATA COLLECTION

Our initial plan was to use the official Instagram API to collect the media data (media content, location coordinates, number of likes and comments, etc.) based on the creation date filter. However, we faced Instagram’s privacy restriction when running the API, as it is meant for building consumer-based applications rather than data crawling. Therefore, we need to find other ways to collect the data we need.

Below are the successful steps we took to collect the Instagram data:

## Step 1. Location IDs: Web Crawling

We found a list of Singapore locations from Instagram’s official website[[3]](#footnote-2), with the location IDs and location names embedded within the raw HTML. This means it is possible to collect all location IDs and names by web-crawling.



Figure 2. HTML content of the list of locations

For example, the highlighted object can be interpreted as:

*Location ID: 42124*

*Name: Pan Pacific Singapore*

## Step 2. Media Data: PHP Crawler

We used a php crawler 6 written by Òscar Casajuana to get some of the media data. The input of the crawler is the location ID, and output is the media data for 12 most recent posts in that locations. The media data we get are the media codes (unique identifier for each posts) and creation time.

## Step 3. Number of Likes: Web Crawling

We looked into the webpage of each media content and realized the number of likes is extractable from the HTML as well. Similar to step 2, we implemented a crawler that visit every post we get from the previous step using its media code, and extracted the number of likes from the web content.

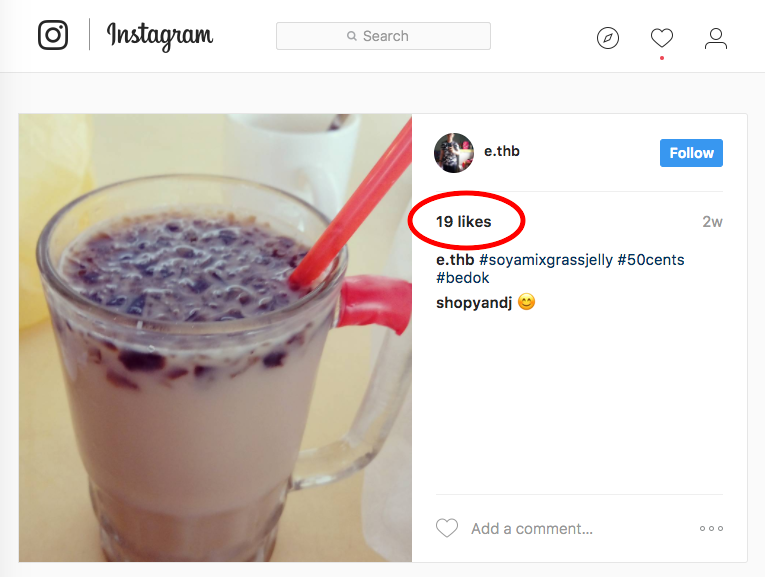


Figure 3. Number of likes on a webpage

## Step 4. Location Latitude & Longitude: Geocoding

We used an online geocoding tool[[4]](#footnote-3) to geocode our location names into latitude and longitude coordinates. The input is a csv file with one field containing all the location names (prepended with “Singapore ” to make geocoding more accurate), and the output is a csv file with 3 fields: name, lat, lon.

Combining the output of all of the steps above, we got the following cleaned up data format which contain around 45,000 entries, with header as shown:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ID | Media code | Date/Time | Number of Likes | Latitude | Longitude | Location Name |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

# DATA ANALYSIS

## Estimate the Number of Posts

As we are unable to count the actual number of posts at one location directly from Instagram API, we need to estimate the count based on post interval.

The csv data file generated from the above steps was first sorted in RStudio in an ascending order of location ID and time. For each post that belong to the same Location ID, the program would calculate the creation time interval between 2 neighbouring post entries. However, exceptional cases happened when either there was only one post belongs to the particular Location ID, or the post was the first for a particular location. For those cases, the time intervals were replaced by the average value of the entire column.

We estimated the number of posts at each location by taking the inverse of the creation time interval. This is based on the assumption that more photos are posted at one location when the creation time interval between posts is short.

## Different Measurement Criteria: Mean vs Median

For each location, we obtained both the mean and the median number of likes (across all 12 posts), the mean and median post creation time interval.

The post density was first estimated by taking the inverse of time interval (both mean and median).

We also estimated the post popularity by calculating the number of likes per post (mean likes \* mean interval, or median likes \* median interval), and plotted the results in QGIS spatial plot.

For better visualization, We did a spatial join between the Hex-Clip map with the vector points of each Instagram post. We select the spatial operation to be ‘contains’ as the points are vectors. Unfortunately, spatial joining of 45000 vectors crashes QGIS, so we merged the points with the same coordinates and consolidate the attributes we need under one point. This leaves us with about 5200 points plotted.

# DATA VISUALIZATION

## Web Interactive Map

From the data analysis step, we generated several maps showing either the post density or popularity, based on either mean or median measurement criteria.

The mainstream approach for comparing different maps is to align them in multi-faceted maps. However, we consider the interactive web interface to be more suitable (with larger viewport area and hyperlinked labels) and more engaging, and thus created the following web-based interactive map:

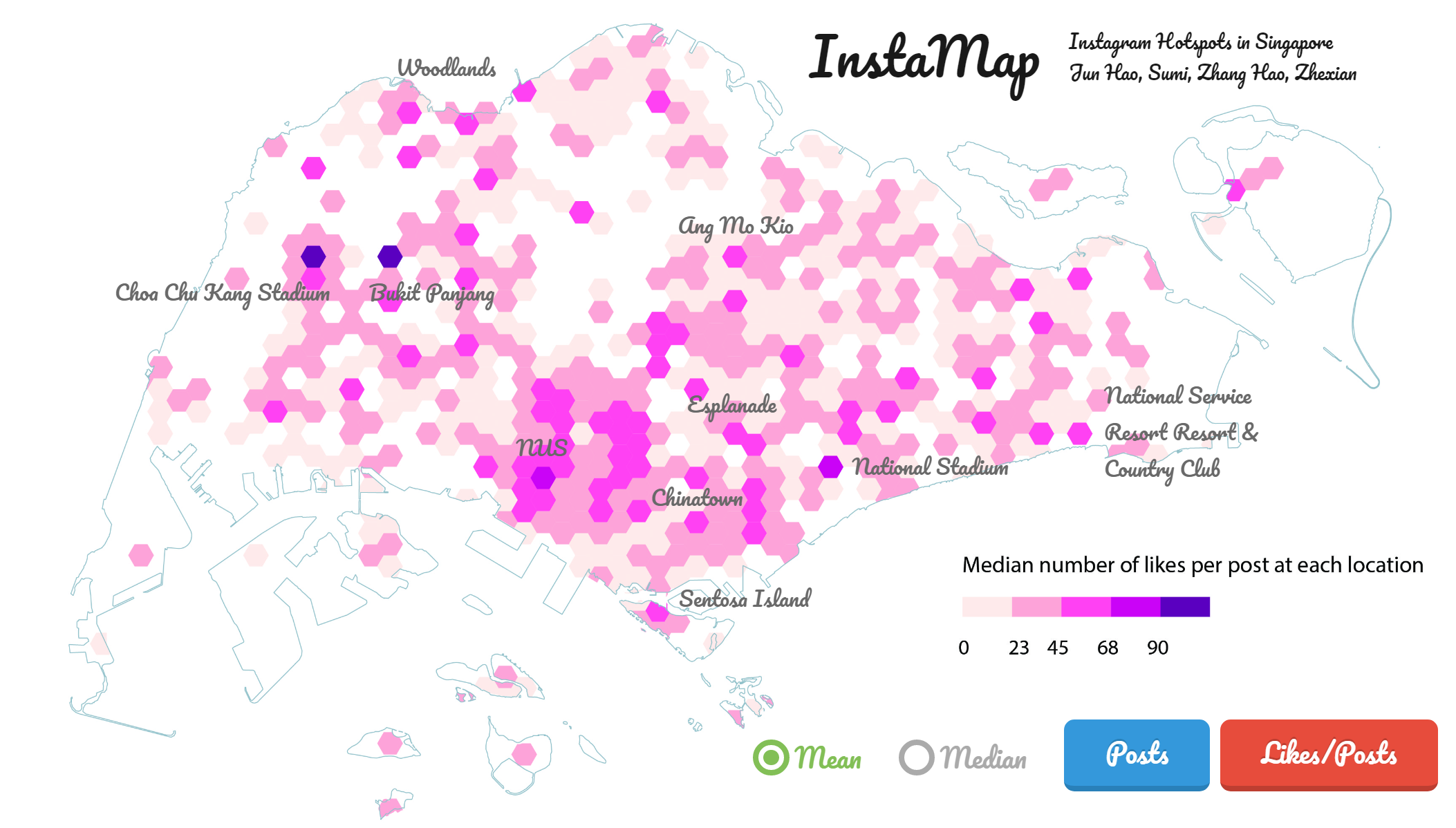


Figure 4. Screenshot of the web interactive map

These are the prominent features of the web map:

* Measurement criteria selection: mean or median
* Topic selection: Posts (density) or Likes/Posts (popularity)
* Hotspot labels with hyperlinks: link to view posts at hotspot location in real time

## 

## 3D Printed Map

The 3D printed map is based on the median likes/posts. We categorized the data into 10 classes in this 3D map to give likes/post higher resolution. This is achieved without compromising the recognisability of each class because map users are more able to tell a unit difference in height than a unit difference in colour. Solidworks 2013 was used to create 3D model of the map. Due to the limitation of a 3D printer’s print area, we have split the map into sectors based on Singapore planning areas.



Figure 5. 3D Printed Map

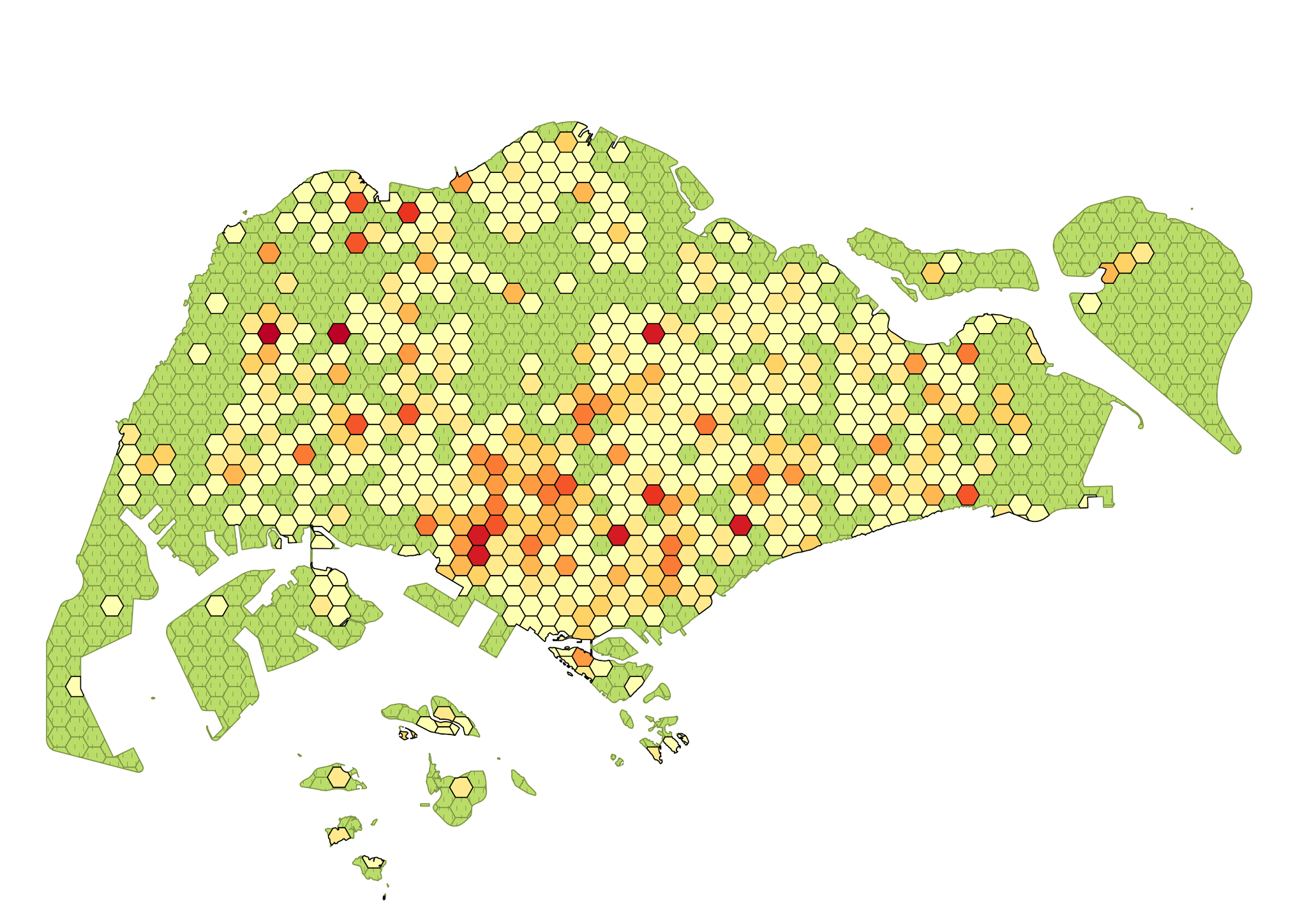


Figure 4. An Image of Median-InstaMap with 10 Classes.

# 

# 

# REFERENCES

1. Instagram usage statistics: <http://connectedlife.tnsglobal.com>
2. Instagram-worthy Places in Singapore: <http://thesmartlocal.com/read/instagram-worthy-places>
3. Instagram locations in Singapore: <https://www.instagram.com/explore/locations/SG/singapore/>
4. Online batch geocoding tool: [www.freegeocoding.com](http://www.freegeocoding.com)
5. InstaMap 3D models: <https://drive.google.com/drive/folders/0B04zZpZF7XEfZDVIRDhMclJKY0U?usp=sharing>
6. PHP Web Crawler: <https://github.com/smochin/instagram-php-crawler>

# APPENDICES

## Appendix A. Potential Benefits of Project (by targeting audience)

* Local: what’s new in hometown, find new and popular photogenic spots
* Tourist: travel like a local, find popular destinations
* Singapore Tourism Board: good way to evaluate popularity of spots and provide guidance for future planning
* Companies: more targeted advertisement at popular spots
* Map Study Students: relatable and engaging topic with sufficient level of complexity in data collection, analysis, and visualization
* Government: measure social media impact and design regulation
* Psychology and Sociology Researchers: provide source of behaviour analysis
* Artist/Photographer: inspiration for photo traveling/representation of Singapore

## Appendix B. Our GitHub Repository

URL: <https://github.com/hellozhanghao/Instamap>

It contains the following files:

* PHP scripts
* Processed maps
* Web interactive map source code
* 3D print file
* R scripts
* Raw data files

## Appendix C Conclusion

We are glad to find the Instagram hotspots in Singapore, which shows that the most popular spots are different from what we expected (Central Business District or tourist attractions). While appreciating the interesting and refreshing findings, we also note the limitations of our project which may be improved in future work:

1. The Instagram media data were not collected at one single time point, as the PHP crawler took more than 20 hours to run
2. The density and popularity data were collected based on estimations: post number was estimated from creation time interval, and the like count was estimated from the likes received by the most recent 12 posts per location
3. Geocoding issues: since instagram allows user to customize location name, we see quite a number of creative/untranslatable names, such as “Bus 165”, “Home sweet home” or “Earth”.

All in all, we enjoyed the process of creating both virtual and physical maps based on a topic close to both our real life and social media life. Hope you like our InstaMap.

1. Infograph: <http://connectedlife.tnsglobal.com> [↑](#footnote-ref-0)
2. Source: <http://thesmartlocal.com/read/instagram-worthy-places> [↑](#footnote-ref-1)
3. <https://www.instagram.com/explore/locations/SG/singapore/>

   6 <https://github.com/smochin/instagram-php-crawler> [↑](#footnote-ref-2)
4. Online batch geocoding tool: [www.freegeocoding.com](http://www.freegeocoding.com/uploadindex.php) [↑](#footnote-ref-3)