MAT259 Proj 2: 2D Spatial Map

Song Gao

Geography Department (Email: sgao@geog.ucsb.edu)

Website: http://www.geog.ucsb.edu/~sgao/

Concept

As a geographer, I am interested in whether I can apply some cartographical design in the 2D spatial maps such as choropleth maps or spatial density maps to visualize the temporal patterns of check-outs in different granularities, such as hourly, daily, weekly and monthly. And more importantly, different visualization methods might have their own pros and cons. I will try different ways to find a good visualization to show the temporal patterns of check-outs. Specifically, a choropleth map is a thematic 2D-map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map. In addition, with regard to the spatial density maps, it calculates the density of features around each output cell (a cell is one or several pixels) based on its value and neighborhoods' values. Conceptually, a neighborhood is defined around each cell center, and the number of cells that fall within the neighborhood is summarized and divided by the area of the neighborhood.

Data and Query

For the X-axis dimension, it is labeled as hourly; for the Y-axis dimension, we can change into different temporal scales, such as days of a week, days of a month, or days of a year.

(1) Hourly check-outs "hot-spots" across a week

```
SELECT dayofweek(cout), hour(cout), count(*)     /*count the number of transactions by day*/

FROM inraw     /*select the data from the target table inraw*/

WHERE date(cout)>='2013-10-07' AND date(cout)<='2013-10-13' /*set the time period*/

GROUP BY dayofweek(cout), hour(cout)
```

ORDER BY dayofweek(cout), hour(cout); /*group the results by different days, note that 1 represents Sunday, 2 for Monday, and such*/

Processing Time: 27.628 seconds and 89 rows returned

Results

http://www.geog.ucsb.edu/~sgao/mat259/2013OctWeekHouly_checkouts.csv

(2) Hourly check-outs "hot-spots" across a month

```
SELECT dayofmonth(cout), hour(cout), count(*)
FROM inraw
WHERE date(cout)>='2013-10-01' AND date(cout)<='2013-10-31'
GROUP BY dayofmonth(cout), hour(cout)
ORDER BY dayofmonth(cout), hour(cout);
```

Processing Time: 27.799 seconds and 373 rows returned

Results

http://www.geog.ucsb.edu/~sgao/mat259/2013OctMonthHouly_checkouts.csv

(3) Hourly check-outs "hot-spots" across a year

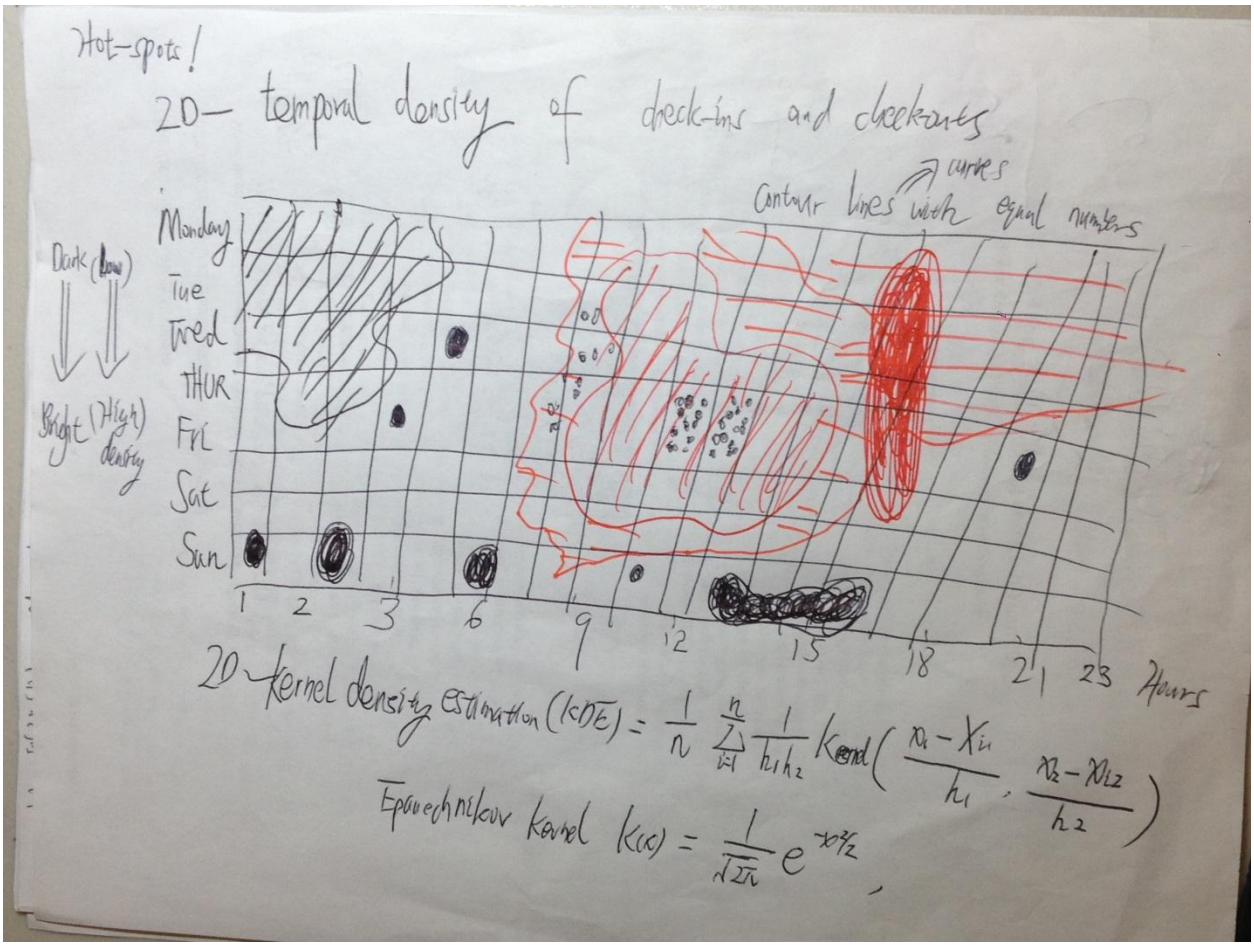
```
SELECT dayofyear(cout), hour(cout), count(*)
FROM inraw
WHERE date(cout)>='2013-01-01' AND date(cout)<='2013-12-31'
GROUP BY dayofyear(cout), hour(cout)
ORDER BY dayofyear(cout), hour(cout);
```

Processing Time: 30.53 seconds and 4319 rows returned

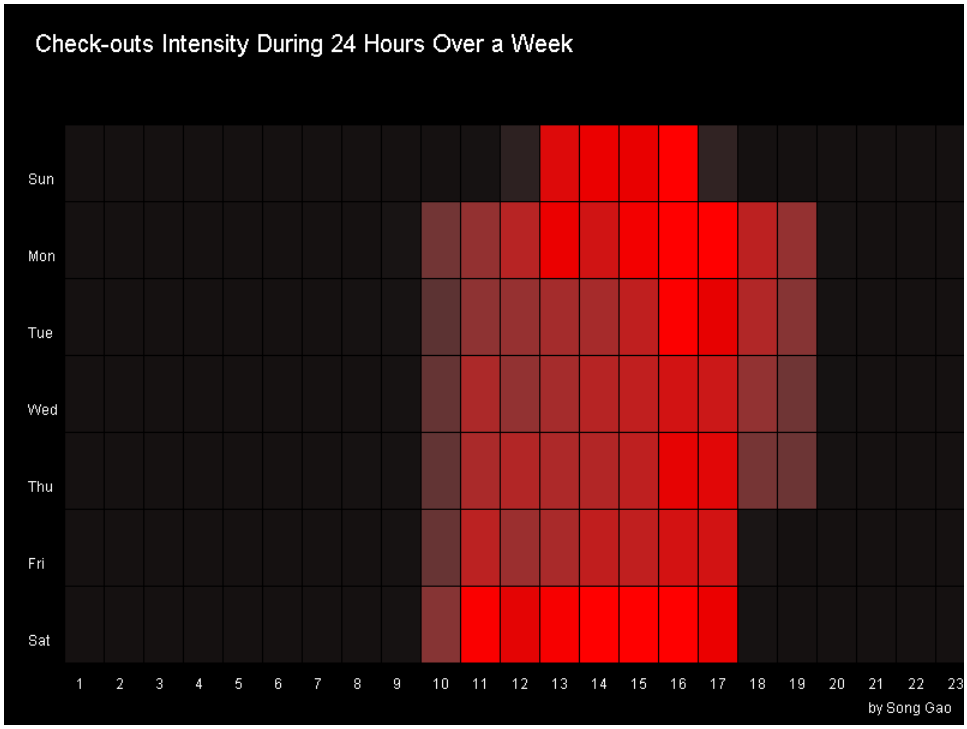
Results

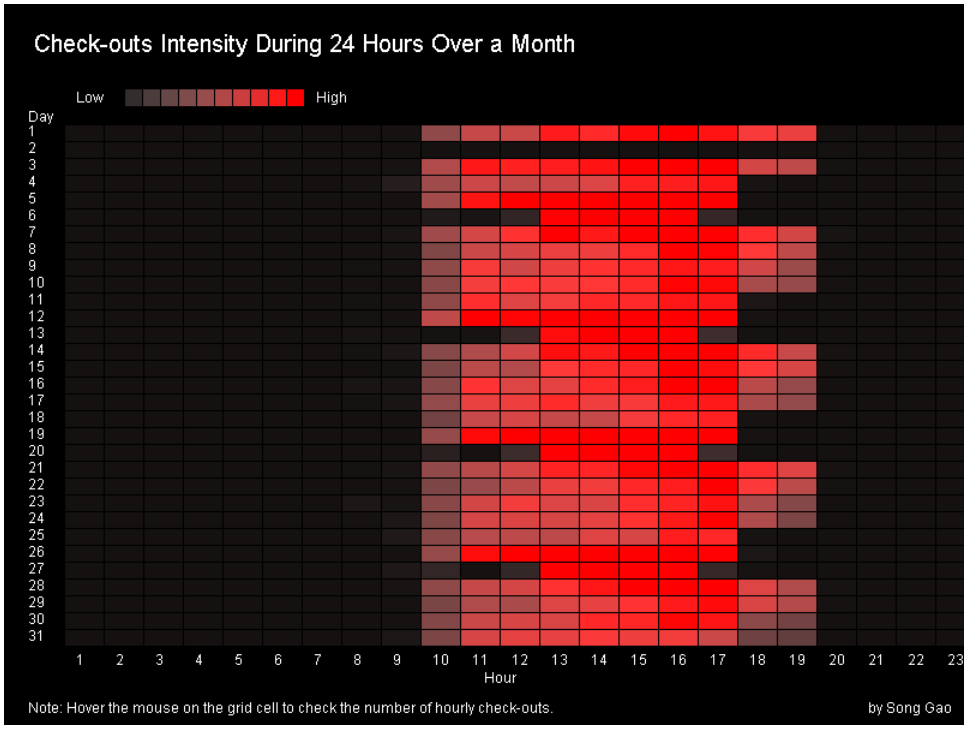
http://www.geog.ucsb.edu/~sgao/mat259/2013YearHouly_checkouts.csv

Doodle:

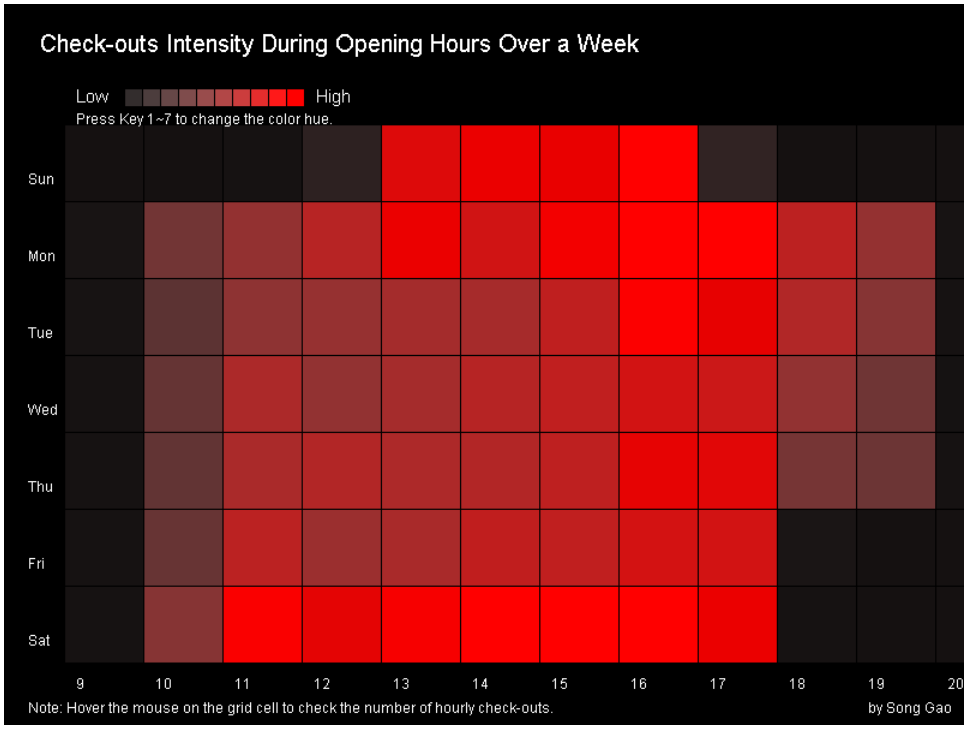

Hot-spots!
2D- temporal density of check-ins and checkouts
Contour lines with equal numbers
curves
Dark (low)
Bright (High) density
Monday
Tue
Wed
THUR
Fri
Sat
Sun
1 2 3 6 9 12 15 18 21 23 Hours
2D-kernel density estimation (KDE) = $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1 h_2}\text{Kernel}\left(\frac{x_1 - x_{i1}}{h_1}, \frac{x_2 - x_{i2}}{h_2}\right)$
Epanechnikov kernel $k(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

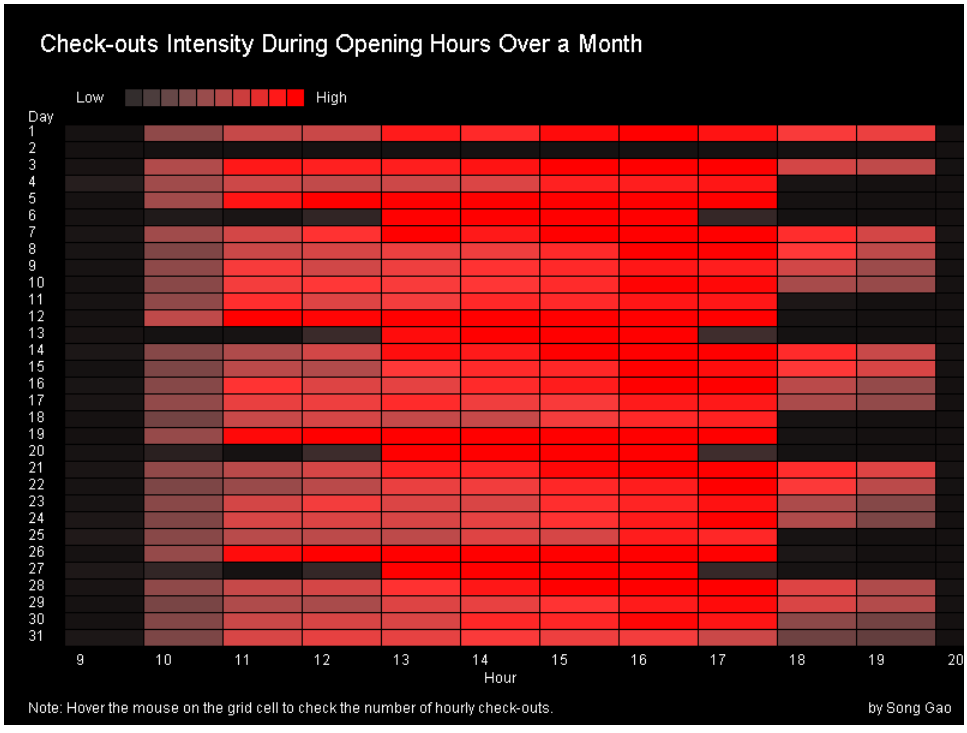
Early Version Visualization:

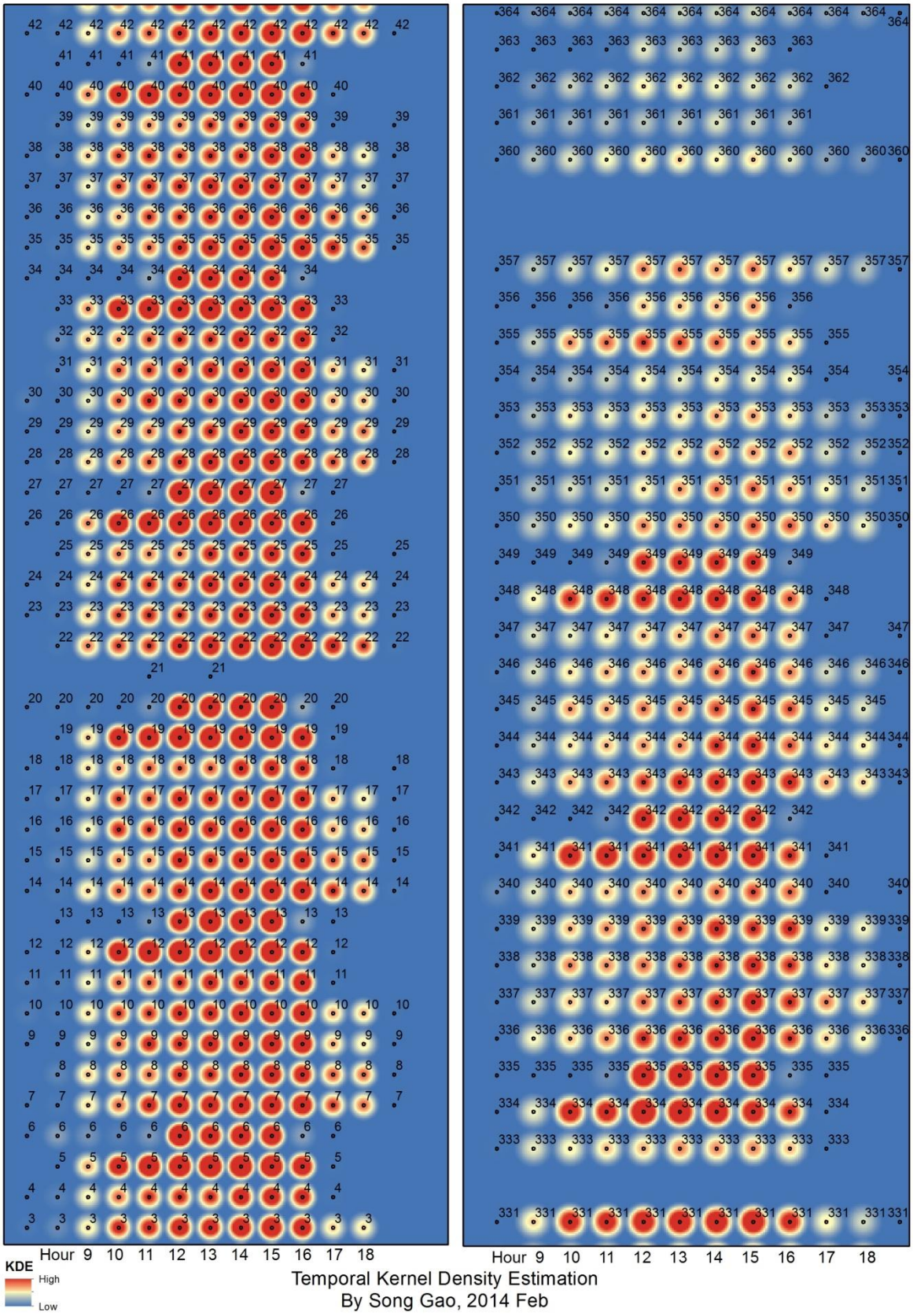


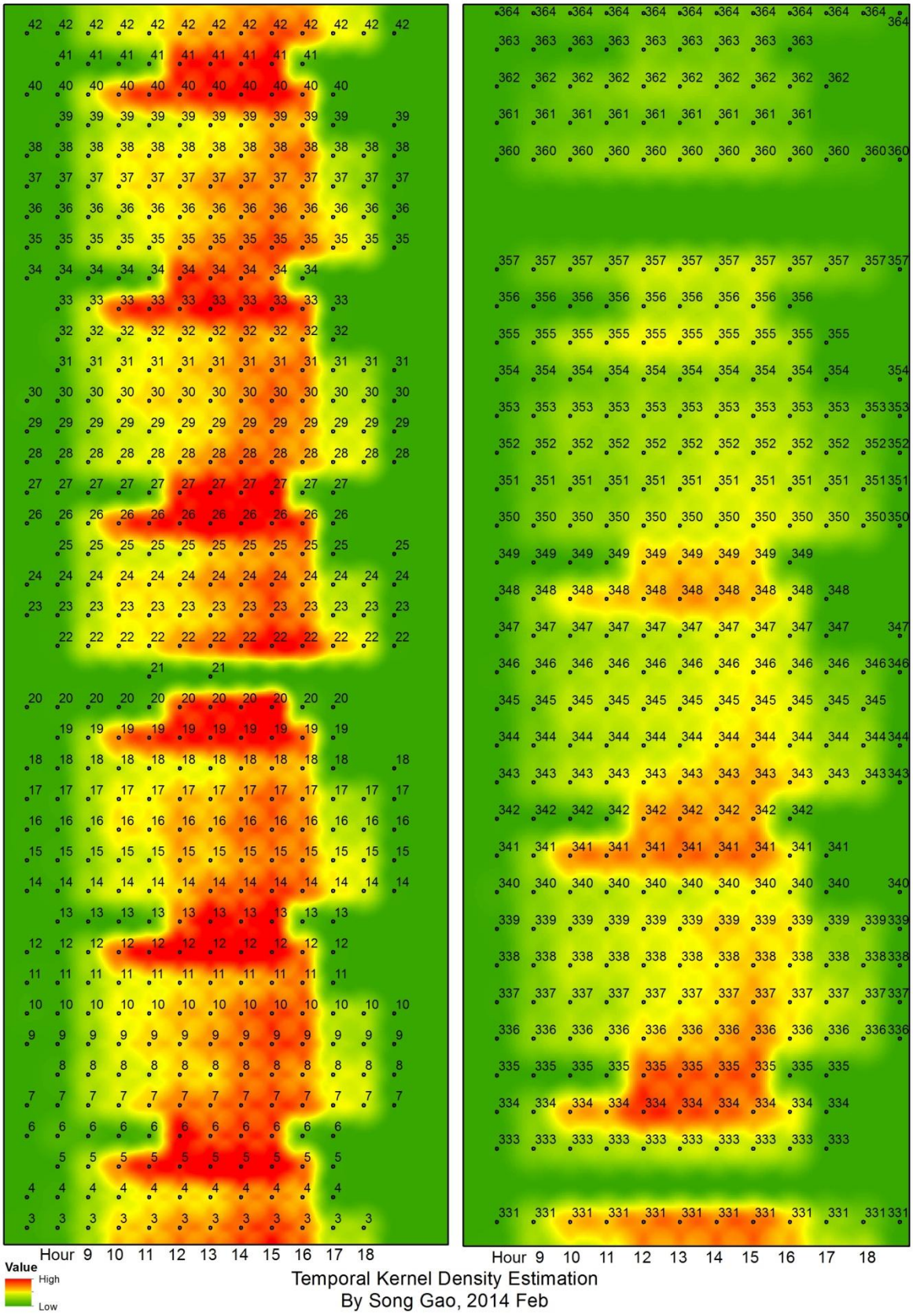


Final Version Visualization:





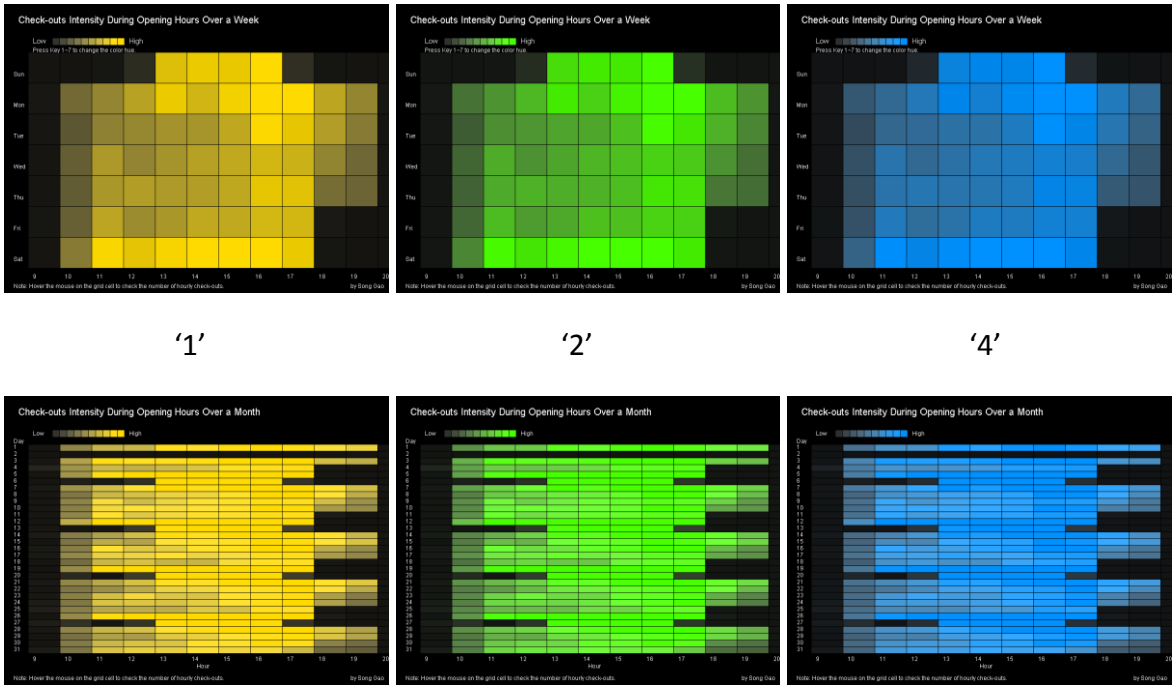

Hour 9 10 11 12 13 14 15 16 17 18
Hour 9 10 11 12 13 14 15 16 17 18
KDE
High
Low
Temporal Kernel Density Estimation
By Song Gao, 2014 Feb


Hour 9 10 11 12 13 14 15 16 17 18
Hour 9 10 11 12 13 14 15 16 17 18
Value
High
Low
Temporal Kernel Density Estimation
By Song Gao, 2014 Feb

Interaction:

The user can hover the mouse on the grid cell to check the actual number of hourly check-outs. One can also press the key 1, 2, 3, 4, 5, 6, 7 to change the color hue.

For example,



'1' '2' '4'

Results and Explanations:

From the Figure "Check-outs Intensity During Opening Hours Over a Week", you can see clearly about the hourly pattern that people tend to return the items between 16~18 in the afternoon and most likely after they finish the daily work. In this specific week, there are also a large number of check-outs on Saturday, from 11AM till 6PM, but for Sunday, such pattern only lasts from 12PM till 5PM, which is related to the Seattle Public Library opening hours on different days and the citizens' behavior preferences. You can find the similar patterns across the month.

More advanced, I used multivariate-kernel-density-estimation (KDE) approach to find the temporal kernel density patterns between hours and days across months. Basically, the KDE generates a smooth surface of the density of check-outs in a temporal neighborhoods around the selected time grid, and the band width (temporal duration) selection in the calculation

affects the results. You can see the different visual results in the temporal-KDE figures (one is [1hour, half-a-day] band width, and the other is [1hour, a-day] band width). It is easy to identify that before the end of this year (during mid-December, day 350~day 364), there are less check-outs compared with the density in January, which has many “red hot-spots” areas. From the hourly perspective, it indicates a higher probability of check-outs between 11AM ~ 5PM.

To know more information about the Kernel Density Estimation if you are interested:

http://en.wikipedia.org/wiki/Kernel_density_estimation

http://en.wikipedia.org/wiki/Multivariate_kernel_density_estimation