experiments-agile

Ming Li

Tuesday, October 28, 2014

#### load the data, get the subset of active users

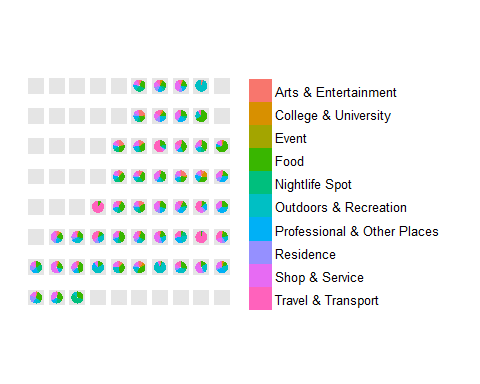
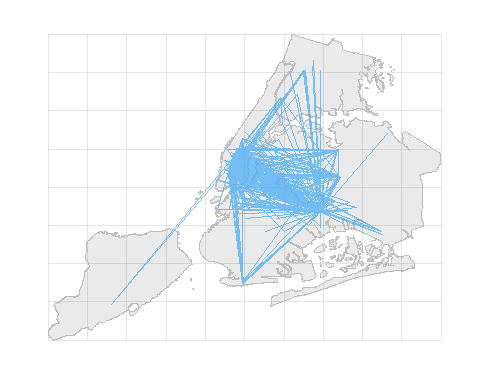
## Loading required package: sp  
## rgdal: version: 0.8-16, (SVN revision 498)  
## Geospatial Data Abstraction Library extensions to R successfully loaded  
## Loaded GDAL runtime: GDAL 1.11.0, released 2014/04/16  
## Path to GDAL shared files: D:/Program Files/R/R-3.1.1/library/rgdal/gdal  
## GDAL does not use iconv for recoding strings.  
## Loaded PROJ.4 runtime: Rel. 4.8.0, 6 March 2012, [PJ\_VERSION: 480]  
## Path to PROJ.4 shared files: D:/Program Files/R/R-3.1.1/library/rgdal/proj  
## Loading required package: grid

#### spatial grid experiment (1)

## OGR data source with driver: ESRI Shapefile   
## Source: "D:\Experiments\foursquare checkin data\shapefile\boundaries", layer: "NYC\_borough\_boundaries\_WGS84"  
## with 5 features and 4 fields  
## Feature type: wkbMultiPolygon with 2 dimensions

## Loading required package: rgeos  
## rgeos version: 0.3-6, (SVN revision 450)  
## GEOS runtime version: 3.4.2-CAPI-1.8.2 r3921   
## Polygon checking: TRUE

## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect  
## Warning: Removed 10 rows containing missing values (position\_stack).  
## Warning: kein nicht-fehlendes Argument für min; gebe Inf zurück  
## Warning: kein nicht-fehlendes Argument für max; gebe -Inf zurück  
## Warning: position\_stack requires constant width: output may be incorrect

* correlation

category.in.grids2 = category.in.grids  
category.in.grids2$uid = factor(category.in.grids2$uid)  
summary(assocstats(xtabs(data=category.in.grids2, ~cate\_l1+uid)))

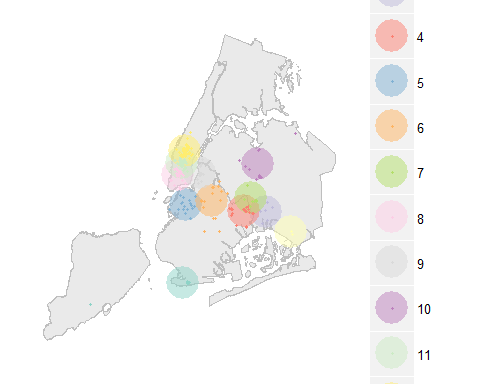
##   
## Call: xtabs(formula = ~cate\_l1 + uid, data = category.in.grids2)  
## Number of cases in table: 578220   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 254324, df = 432, p-value = 0  
## Chi-squared approximation may be incorrect  
## X^2 df P(> X^2)  
## Likelihood Ratio 171268 432 0  
## Pearson 254324 432 0  
##   
## Phi-Coefficient : 0.663   
## Contingency Coeff.: 0.553   
## Cramer's V : 0.221

summary(assocstats(xtabs(data=category.in.grids2, ~cate\_l2+uid)))

##   
## Call: xtabs(formula = ~cate\_l2 + uid, data = category.in.grids2)  
## Number of cases in table: 578220   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 1449908, df = 18336, p-value = 0  
## Chi-squared approximation may be incorrect  
## X^2 df P(> X^2)  
## Likelihood Ratio 564458 18336 0  
## Pearson 1449908 18336 0  
##   
## Phi-Coefficient : 1.584   
## Contingency Coeff.: 0.846   
## Cramer's V : 0.229

#### spatial grid experiment (2): for single user, in cluster

## Warning: n too large, allowed maximum for palette Set3 is 12  
## Returning the palette you asked for with that many colors  
##   
## Warning: n too large, allowed maximum for palette Set3 is 12  
## Returning the palette you asked for with that many colors  
##   
## Warning: n too large, allowed maximum for palette Set3 is 12  
## Returning the palette you asked for with that many colors



* correlation

summary(assocstats(xtabs(~cate\_l1+cid.ordered, data=category.in.clusters$point)))

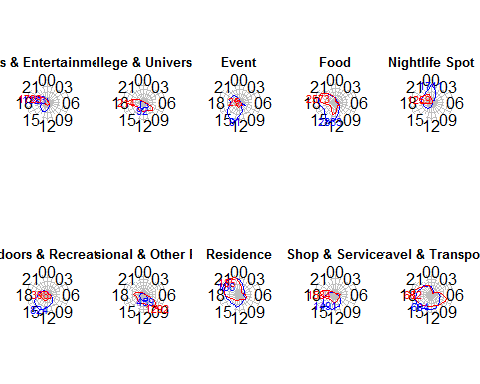
##   
## Call: xtabs(formula = ~cate\_l1 + cid.ordered, data = category.in.clusters$point)  
## Number of cases in table: 3580   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 3131, df = 117, p-value = 0  
## Chi-squared approximation may be incorrect  
## X^2 df P(> X^2)  
## Likelihood Ratio 1655.2 117 0  
## Pearson 3130.9 117 0  
##   
## Phi-Coefficient : 0.935   
## Contingency Coeff.: 0.683   
## Cramer's V : 0.312

summary(assocstats(xtabs(~cate\_l2+cid.ordered, data=category.in.clusters$point)))

##   
## Call: xtabs(formula = ~cate\_l2 + cid.ordered, data = category.in.clusters$point)  
## Number of cases in table: 3580   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 18180, df = 1339, p-value = 0  
## Chi-squared approximation may be incorrect  
## X^2 df P(> X^2)  
## Likelihood Ratio 5858.1 1339 0  
## Pearson 18179.7 1339 0  
##   
## Phi-Coefficient : 2.253   
## Contingency Coeff.: 0.914   
## Cramer's V : 0.625

# temporal experiment

##   
## Attaching package: 'plotrix'  
##   
## The following object is masked from 'package:scales':  
##   
## rescale



* correlation

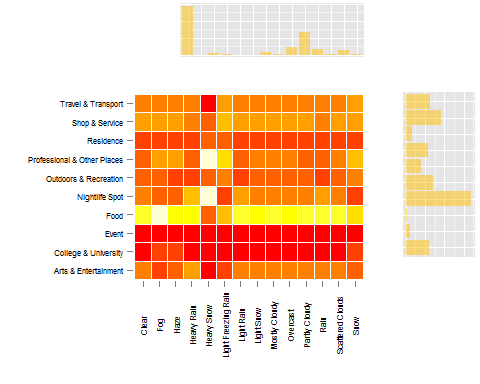
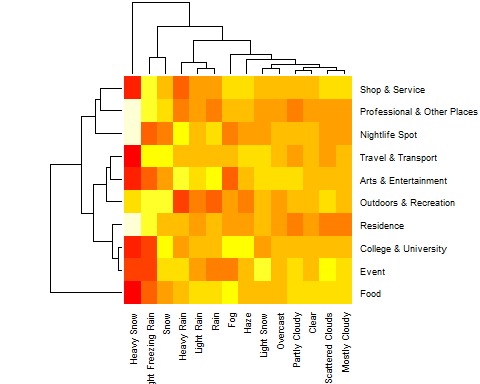
summary(assocstats(xtabs(~cate\_l1 + hour, data = checkin.global)))

##   
## Call: xtabs(formula = ~cate\_l1 + hour, data = checkin.global)  
## Number of cases in table: 578220   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 143030, df = 207, p-value = 0  
## X^2 df P(> X^2)  
## Likelihood Ratio 132875 207 0  
## Pearson 143030 207 0  
##   
## Phi-Coefficient : 0.497   
## Contingency Coeff.: 0.445   
## Cramer's V : 0.166

summary(assocstats(xtabs(~cate\_l2 + hour, data = checkin.global)))

##   
## Call: xtabs(formula = ~cate\_l2 + hour, data = checkin.global)  
## Number of cases in table: 578220   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 288656, df = 8786, p-value = 0  
## Chi-squared approximation may be incorrect  
## X^2 df P(> X^2)  
## Likelihood Ratio 255802 8786 0  
## Pearson 288656 8786 0  
##   
## Phi-Coefficient : 0.707   
## Contingency Coeff.: 0.577   
## Cramer's V : 0.147

#### weather impact experiment

* correlation

summary(assocstats(xtabs(~cate\_l1 + conds, data = checkin.global)))

##   
## Call: xtabs(formula = ~cate\_l1 + conds, data = checkin.global)  
## Number of cases in table: 578220   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 3591, df = 117, p-value = 0  
## Chi-squared approximation may be incorrect  
## X^2 df P(> X^2)  
## Likelihood Ratio 3570.7 117 0  
## Pearson 3590.8 117 0  
##   
## Phi-Coefficient : 0.079   
## Contingency Coeff.: 0.079   
## Cramer's V : 0.026

summary(assocstats(xtabs(~cate\_l2 + conds, data = checkin.global)))

##   
## Call: xtabs(formula = ~cate\_l2 + conds, data = checkin.global)  
## Number of cases in table: 578220   
## Number of factors: 2   
## Test for independence of all factors:  
## Chisq = 15572, df = 4966, p-value = 0  
## Chi-squared approximation may be incorrect  
## X^2 df P(> X^2)  
## Likelihood Ratio 13149 4966 0  
## Pearson 15572 4966 0  
##   
## Phi-Coefficient : 0.164   
## Contingency Coeff.: 0.162   
## Cramer's V : 0.046