Lab 05 - Heat Maps

Jesse Lecy

Wednesday, February 18, 2015

This lab introduces heat maps (sometimes known as contour plots) as a way to summarize dense points on a map.

We will use the *stat_density2d* function in the *ggmap* package to add points to our plot. It requires the following arguments that are identical to the *geom_points* function arguments:

Argument	Values
data=	Name of the data set that contains the latitude and longitude coordinates.
aes=	The names of your latitude and longitude variables in the data set.
size=	The size of your points with 1 being default, 2 being double that, etc.
col=	The color of your points. For options type *colors()*.
alpha=	THe level of transparency of your points. Accepts values between 0 (invisible) and 1 (solid).

And three additional arguments:

Argument	Values
bins=	Controls the granularity of aggregation of your data.
geom=	Either *polygon* or *density2d* to set the aesthetic.

311 Service Call Data

For this lab we will be using 311 Service Call data from the NYC Open Data project. We are specifically using all 311 calls from 2009 available here:

https://data.cityofnewyork.us/Social-Services/311-Service-Requests-2009/3rfa-3xsf

I have downloaded the data set and posted it as a file on BlackBoard called "311 Service Requests for 2009.csv".

```
# To download from the NYC Open Data site:

# library( RCurl ) package to read data from html files
#
# copy the download link by right-clicking on the 'download as csv' link
#
# url <- "http://data.cityofnewyork.us/api/views/3rfa-3xsf/rows.csv?accessType=DOWNLOAD"
#
# service.calls <- read.csv( file=url, stringsAsFactors=F )
#
# dat.taxi <- service.calls[ service.calls$Complaint.Type == "Taxi Complaint" , ]
```

```
setwd("C:/Users/jdlecy/Documents")

dat <- read.csv( "311 Service Requests for 2009.csv", stringsAsFactors=F )

dim( dat )  # should have 1,783,133 observations and 52 variables
```

```
## [1] 1783133      52
```

```
names( dat )
```

```
##  [1] "Unique.Key"                    "Created.Date"
##  [3] "Closed.Date"                   "Agency"
##  [5] "Agency.Name"                   "Complaint.Type"
##  [7] "Descriptor"                    "Location.Type"
##  [9] "Incident.Zip"                  "Incident.Address"
## [11] "Street.Name"                   "Cross.Street.1"
## [13] "Cross.Street.2"                "Intersection.Street.1"
## [15] "Intersection.Street.2"         "Address.Type"
## [17] "City"                          "Landmark"
## [19] "Facility.Type"                 "Status"
## [21] "Due.Date"                      "Resolution.Action.Updated.Date"
## [23] "Community.Board"               "Borough"
## [25] "X.Coordinate..State.Plane."    "Y.Coordinate..State.Plane."
## [27] "Park.Facility.Name"            "Park.Borough"
## [29] "School.Name"                   "School.Number"
## [31] "School.Region"                 "School.Code"
## [33] "School.Phone.Number"           "School.Address"
## [35] "School.City"                   "School.State"
## [37] "School.Zip"                    "School.Not.Found"
## [39] "School.or.Citywide.Complaint"  "Vehicle.Type"
## [41] "Taxi.Company.Borough"          "Taxi.Pick.Up.Location"
## [43] "Bridge.Highway.Name"           "Bridge.Highway.Direction"
## [45] "Road.Ramp"                     "Bridge.Highway.Segment"
## [47] "Garage.Lot.Name"               "Ferry.Direction"
## [49] "Ferry.Terminal.Name"           "Latitude"
## [51] "Longitude"                     "Location"
```

Note that the 311 Service Calls data already comes with a Latitude and Longitude variable, so we do not need to geocode the incidents (which is good since there are over 1.7 million observations and we can only geocode 2,500 a day!). Note that the variable names for lat and lon are different than last week.

For this lab we will be looking at various types of complaints within the city to better understand the geographic distribution of specific issues.

Let's take a look at the common types of complaints in our data.

```
as.data.frame( sort(table( dat$Complaint.Type ), decreasing=T ))
```

```
##                                    sort(table(dat$Complaint.Type), decreasing = T)
## HEATING                                                                     238995
## Street Light Condition                                                      130242
## GENERAL CONSTRUCTION                                                        121048
## PLUMBING                                                                    104931
```

```
## Street Condition                                                102150
## PAINT - PLASTER                                                  84324
## NONCONST                                                         67325
## Water System                                                     62991
## General Construction/Plumbing                                    61288
## Traffic Signal Condition                                         60995
## Blocked Driveway                                                 52163
## Sewer                                                            49849
## Dirty Conditions                                                 40732
## ELECTRIC                                                         38349
## Building/Use                                                     35495
## Noise                                                            33707
## Sanitation Condition                                             32851
## Noise - Street/Sidewalk                                          27923
## Noise - Commercial                                               25234
## Rodent                                                           21622
## Taxi Complaint                                                   20909
## Noise - Vehicle                                                  20237
## Damaged Tree                                                     19738
## Consumer Complaint                                               18935
## Graffiti                                                         17947
## Missed Collection (All Materials)                                17355
## Derelict Vehicle                                                 15648
## APPLIANCE                                                        15385
## Overgrown Tree/Branches                                          15201
## Broken Muni Meter                                                12320
## Derelict Vehicles                                                11671
## Snow                                                             11432
## Dead Tree                                                         9885
## Elevator                                                          9098
## Air Quality                                                       8401
## DCA / DOH New License Application Request                         8364
## Maintenance or Facility                                           8241
## Root/Sewer/Sidewalk Condition                                     8205
## Sidewalk Condition                                                6785
## Food Establishment                                                6543
## Street Sign - Damaged                                             5931
## Indoor Air Quality                                                5897
## Other Enforcement                                                 5300
## Broken Parking Meter                                              5283
## Vending                                                           4877
## Electrical                                                        4469
## Plumbing                                                          3823
## Water Conservation                                                3770
## Street Sign - Missing                                             3537
## Emergency Response Team (ERT)                                     3121
## Noise - Park                                                      3038
## Litter Basket / Request                                           2917
## Highway Condition                                                 2893
## Special Projects Inspection Team (SPIT)                           2823
## Food Poisoning                                                    2581
## Hazardous Materials                                               2525
## Vacant Lot                                                        2500
## Homeless Encampment                                               2336
```

```
## For Hire Vehicle Complaint                          2327
## Violation of Park Rules                              2207
## Boilers                                              2198
## Smoking                                              2168
## Unsanitary Animal Pvt Property                       2098
## Lead                                                 1971
## Disorderly Youth                                     1926
## Industrial Waste                                     1841
## BEST/Site Safety                                     1759
## Special Enforcement                                  1752
## Overflowing Litter Baskets                           1738
## Investigations and Discipline (IAD)                  1732
## Sweeping/Missed-Inadequate                           1730
## Animal in a Park                                     1706
## Asbestos                                             1632
## Drinking                                             1620
## Public Payphone Complaint                            1435
## Non-Residential Heat                                 1388
## Recycling Enforcement                                1371
## Street Sign - Dangling                               1363
## Noise - House of Worship                             1349
## Water Quality                                        1349
## Illegal Tree Damage                                  1334
## CONSTRUCTION                                         1186
## DOE Complaint or Compliment                          1057
## DOF Property - Payment Issue                         1003
## Window Guard                                          990
## Safety                                                982
## School Maintenance                                    977
## Cranes and Derricks                                   787
## Unsanitary Pigeon Condition                           753
## Unleashed Dog                                         740
## Discipline and Suspension                             688
## Found Property                                        643
## Teaching/Learning/Instruction                         635
## Health                                                613
## Taxi Compliment                                       591
## Bridge Condition                                      557
## Posting Advertisement                                 507
## Ferry Inquiry                                         475
## DOF Property - City Rebate                            463
## City Vehicle Placard Complaint                        460
## Scaffold Safety                                       415
## Bike/Roller/Skate Chronic                             404
## Registration and Transfers                            392
## Miscellaneous Categories                              389
## DOF Property - Owner Issue                            372
## Urinating in Public                                   359
## Bus Stop Shelter Placement                            355
## Senior Center Complaint                               350
## Tree Work Permit Request                              348
## Ferry Complaint                                       336
## Plant                                                 301
## Illegal Fireworks                                     244
```

```
## Illegal Animal Kept as Pet                                  236
## Collection Truck Noise                                      233
## Panhandling                                                 231
## Beach/Pool/Sauna Complaint                                  184
## Harboring Bees/Wasps                                        163
## Mold                                                        144
## Portable Toilet                                             112
## Summer Camp                                                  97
## Tattooing                                                    87
## Municipal Parking Facility                                   83
## Poison Ivy                                                   81
## DOF Parking - Payment Issue                                  77
## Highway Sign - Damaged                                       77
## Parent Leadership                                            71
## Special Natural Area District (SNAD)                         71
## Unsanitary Animal Facility                                   70
## Animal Facility - No Permit                                  64
## Stalled Sites                                                62
## Drinking Water                                               57
## No Child Left Behind                                         48
## Illegal Animal Sold                                          38
## Calorie Labeling                                             37
## Legal Services Provider Complaint                            34
## Public Toilet                                                31
## Ferry Permit                                                 28
## Bottled Water                                                24
## Illegal Animal - Sold/Kept                                   23
## Highway Sign - Missing                                       21
## X-Ray Machine/Equipment                                      19
## DOF Property - RPIE Issue                                    16
## Lifeguard                                                    16
## Highway Sign - Dangling                                      15
## Parking Card                                                 13
## Adopt-A-Basket                                               12
## Radioactive Material                                         11
## Trans Fat                                                    11
## Tunnel Condition                                             11
## Transportation Provider Complaint                             9
## Squeegee                                                      8
## Forensic Engineering                                          6
## Trapping Pigeon                                               1
```

Point Map of Taxi Cab Complaints

Let us start by creating a map of all of the taxi complaints in the city. I have already created a subset of the data for taxi complaints and saved it to a separate file in case you would like to follow along.

```
# dat.taxi <- service.calls[ service.calls$Complaint.Type == "Taxi Complaint" , ]
#
# write.csv( dat.taxi, "311 Calls for Taxi Complaint.csv", row.names=F )

dat.taxi <- read.csv( "311 Calls for Taxi Complaint.csv", stringsAsFactors=F )
```

```
dim( dat.taxi )  # should have 20,909 observations and 52 variables
```

```
## [1] 20909    52
```

Since we have the latitude and longitude coordinates we can proceed by mappint points in the usual fashion. I find that grabbing a base map that is centered at Queens gives a better view of all boroughs versus one centered at Manhattan.

```
library( ggmap )
```

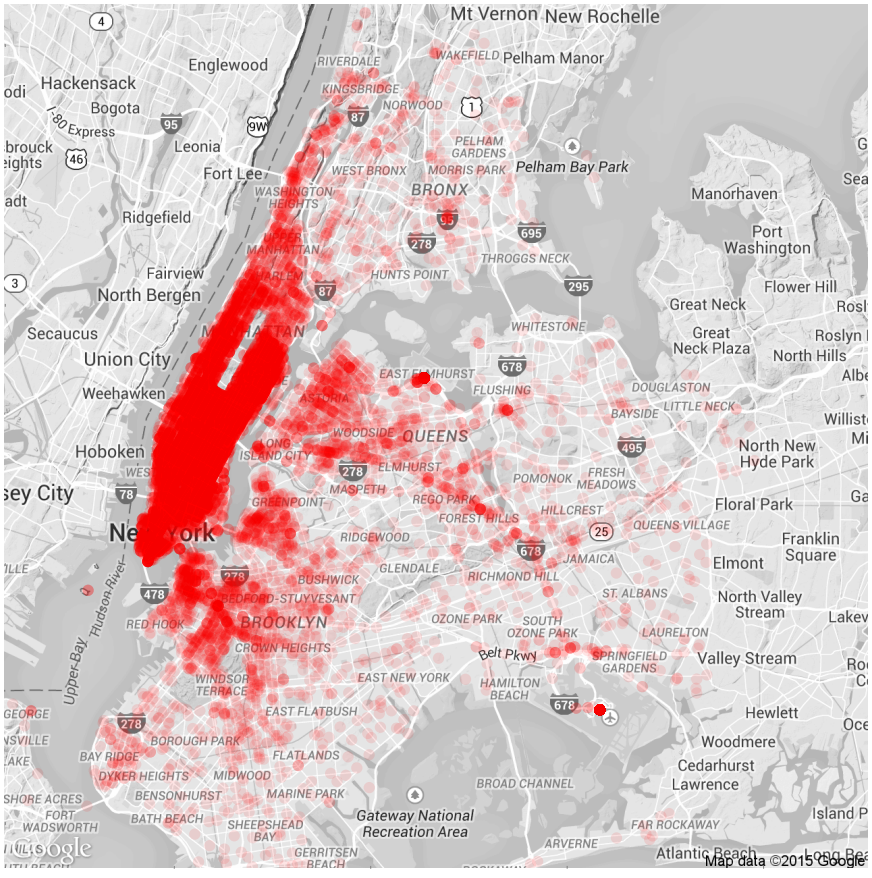
```
## Loading required package: ggplot2
```

```
nyc <- ggmap( get_map( "queens, ny", zoom=11, color="bw" ), extent="device" )
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=queens,+ny&zoom=11&size=%20640x6
## Google Maps API Terms of Service : http://developers.google.com/maps/terms
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=queens,+ny&sensor=fa
## Google Maps API Terms of Service : http://developers.google.com/maps/terms
```

```
nyc + geom_point( data=dat.taxi, aes(x=Longitude, y=Latitude), size=2, col="red", alpha=0.1 )
```

```
## Warning: Removed 1694 rows containing missing values (geom_point).
```



Note that there are warnings produced for missing values. This can mean either that the latitude and longitude data was not available for specific observations (perhaps a complaint was lodged but it did not occure in a specific geography), or else the point is located off of the current base map.

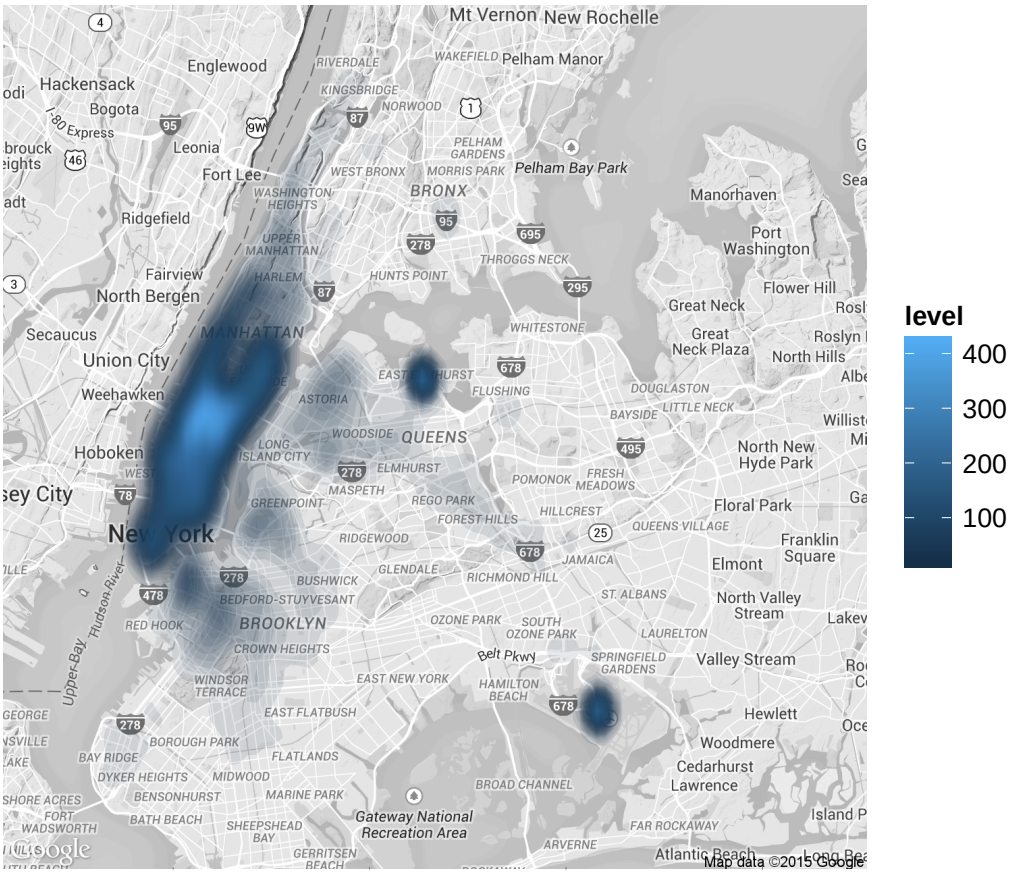
Points Represented as Density - The Heat Map

The mapping functions worked as expected but we have a problem. With over 20,000 points on the map our data is very dense. It is hard to highlight any specific take-aways other than a lot of people complain about taxis, and mostly in Manhattan. Let's see if we can improve upon this.

```
nyc2 <- nyc + stat_density2d( data=dat.taxi, aes(x=Longitude, y=Latitude, fill=..level..), alpha=0.1,
                              bins=100, geom="polygon" )

nyc2
```

```
## Warning: Removed 1694 rows containing non-finite values (stat_density2d).
```



Now we have something interesting to work with. We can see that the point data has now been summarized by the intensity of points per geographic unit. The colors represent density - the average number of points per unit space over a region. We can see where events are very dense, and where they are sparse.

This is particularly useful when point data is dense - many points are overlaid on top of each other (like the figure above). As a result of the overlay, we cannot discern the concentration of specific points.

An example of this emerges from the taxi complaints at the airport in the bottom-right quadrant. On the point map the data does not look that significant because it is all concentrated in a small space so it looks like many other points on the map, but once translated into density we can see that the concentration is significant. Similarly, we see that taxi complaints are most intense in Midtown Manhattan, although there are complaints spread across the whole island.

Layering Data

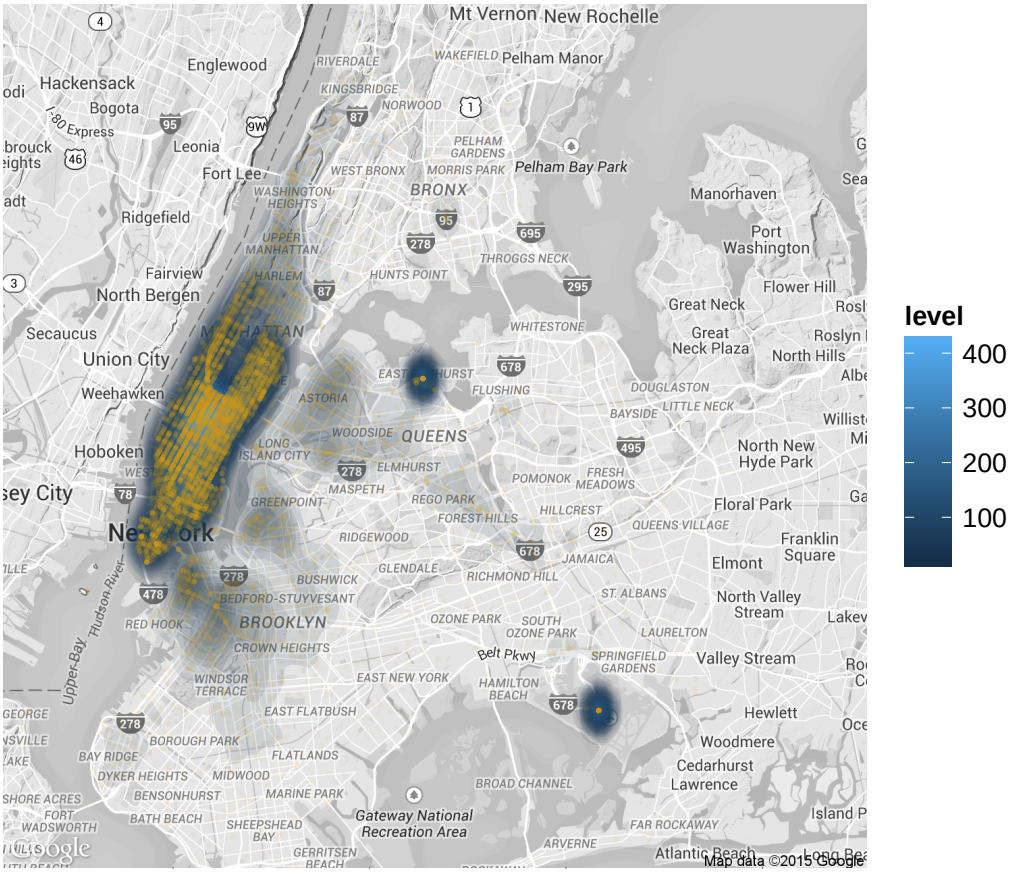
We can enhance the density plots by adding another layer to combine the original point data with the densities. We want to set the alpha high so that we can see specific concentrations of points, but we still see the underlying density structure clearly. I have also chosed a yellow hue that compliments the black and blue colors in the density map.

```
nyc3 <- nyc2 + geom_point( data=dat.taxi, aes(x=Longitude, y=Latitude), size=1, col="goldenrod", alpha=0

nyc3
```

```
## Warning: Removed 1694 rows containing non-finite values (stat_density2d).
```

```
## Warning: Removed 1694 rows containing missing values (geom_point).
```

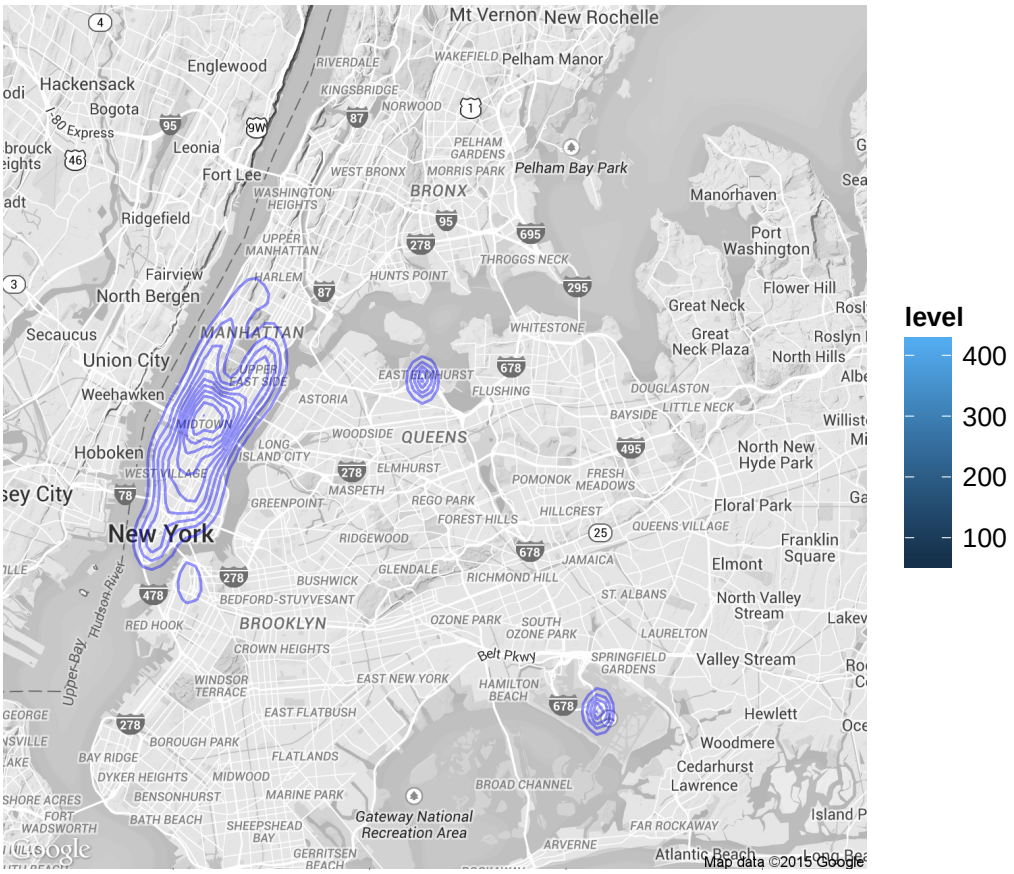


Layering data like this helps us pick up on interesting patterns. For example, we can discern the grid structure of the city based upon locations of complaints, and Central Park is also discernable. More data is not always better (we hide some of the density gradiant now), but layering can be a useful visual tool.

This example above demonstrates the use of the "polygon" value in the geom argument. It produces a gradiant where the density plot is represented by smooth transitions between levels. Alternatively, we can use a contour plot representation of the levels using the "density2d" value for the geom argument:

```
nyc + stat_density2d( data=dat.taxi, aes(x=Longitude, y=Latitude, fill=..level..),
                              bins=10, geom="density2d", alpha=0.4, col="blue" )
```

```
## Warning: Removed 1694 rows containing non-finite values (stat_density2d).
```



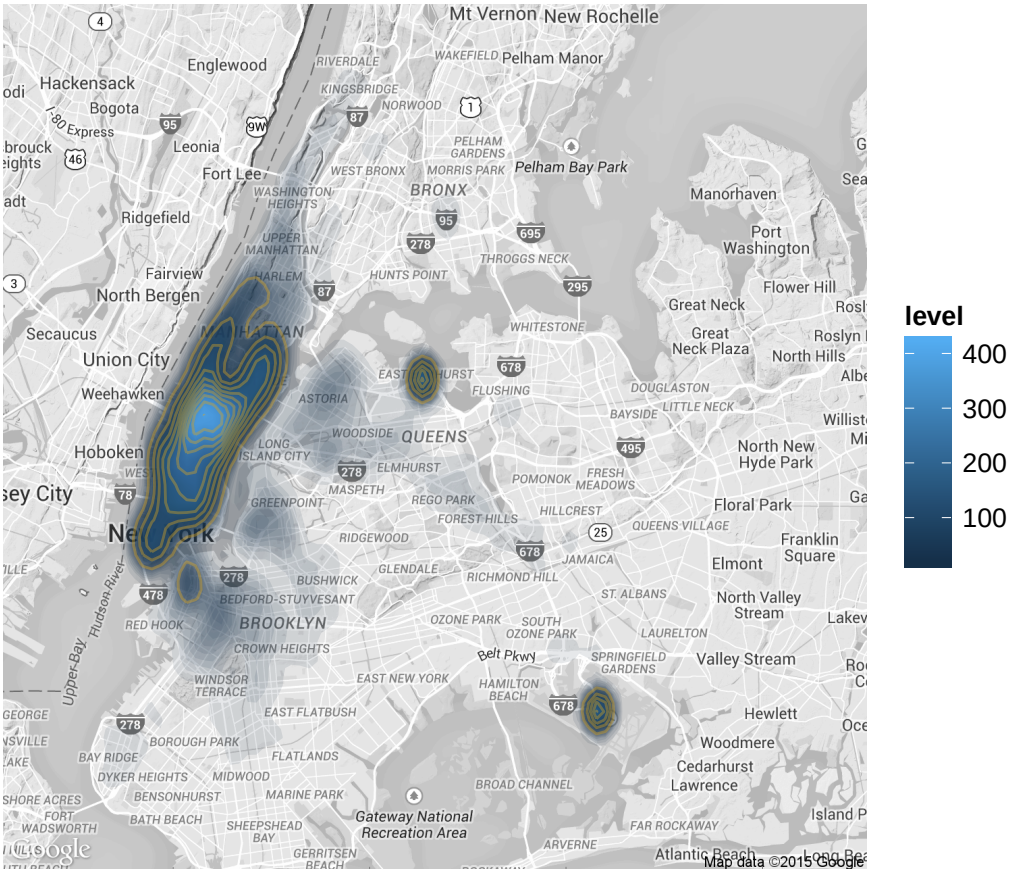
We can also get creative and layer the contour plot on top of the gradiant heat map to highlight regions in the data.

```
nyc3 <- nyc2 + stat_density2d( data=dat.taxi, aes(x=Longitude, y=Latitude, fill=..level..),
                               bins=10, geom="density2d", alpha=0.4, col="goldenrod" )

nyc3
```

```
## Warning: Removed 1694 rows containing non-finite values (stat_density2d).
```

```
## Warning: Removed 1694 rows containing non-finite values (stat_density2d).
```



In summary, density plots are very useful for describing the concentration of points on a map. It creates a clear representation of the intensity of events over a geography.

Looking Ahead

Next week we will begin working with another convention - aggregating data by administrative units. We will use choropleth maps to represent things like levels of household income in each census tract, or crime rates in each neighborhood. This will require us to use shape files, additional map elements for the administrative boundaries within cities.