sample-analysis

Julian DeGroot-Lutzner & Adi Salwan

12/13/2017

f

```
library(readr)
randomsample <- read_csv("~/Documents/math154/ma154-project24-teambike/final_project/randomsample.csv")
```

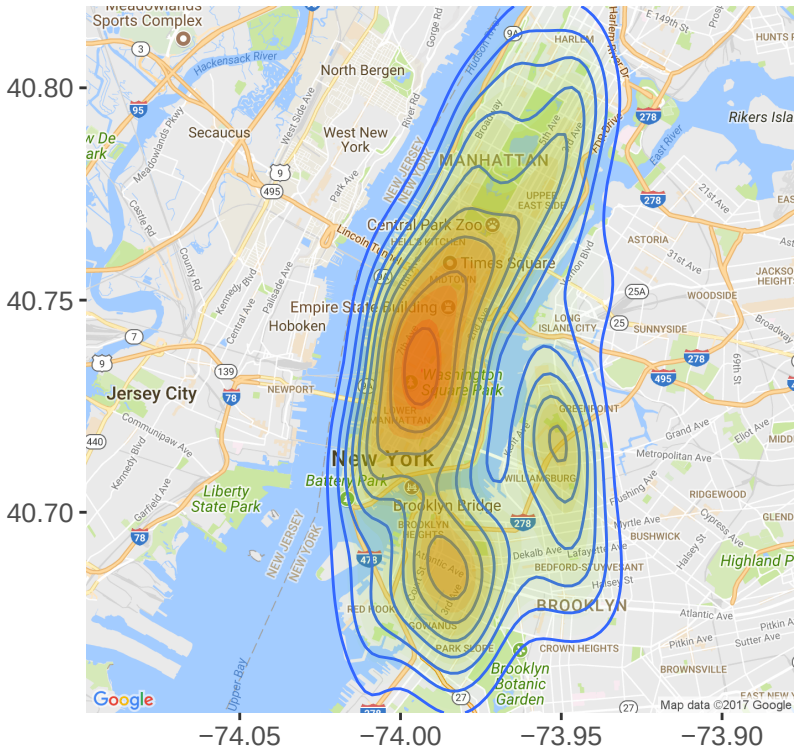
```
start_sums <- randomsample %>%
  group_by(start.station.id) %>%
  summarize(start.station.longitude = mean(start.station.longitude),
            start.station.latitude = mean(start.station.latitude),
            total.time.out = sum(tripduration),
            start.count = n()) %>%
  mutate(avg.time.out = total.time.out/start.count) %>%
  select(-total.time.out) %>%
  ungroup()
```

The Map Used

```
center.citibikes <- c(
  lon = mean(randomsample$start.station.longitude),
  lat = mean(randomsample$start.station.latitude))
mymap <- get_map(location = center.citibikes,
                 maptype = "roadmap",
                 zoom = 12)
```

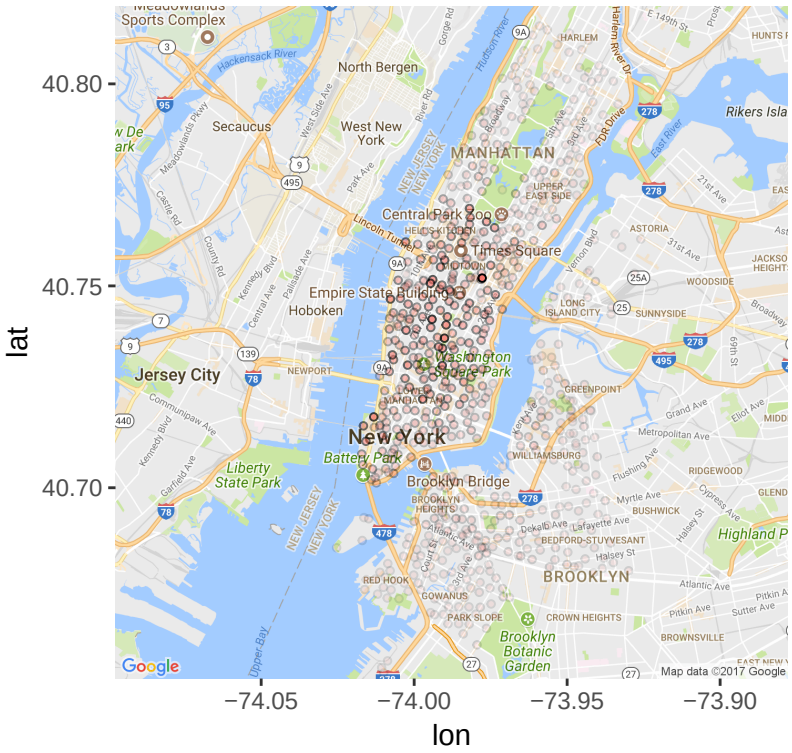
Location of Stations Map

```
 ggmap(mymap, extent = "panel", maprange=FALSE) +
  geom_density2d(data = start_sums,
                           aes(x = start.station.longitude,
                               y = start.station.latitude)) +
  stat_density2d(data = start_sums,
                           aes(x = start.station.longitude,
                               y = start.station.latitude, fill = ..level.., alpha = ..level..),
                 size = 0.01, bins = 16, geom = 'polygon') +
  scale_fill_gradient(low = "green", high = "red") +
  scale_alpha(range = c(0.00, 0.25), guide = FALSE) +
  theme(legend.position = "none", axis.title = element_blank(), text = element_text(size = 12))
```



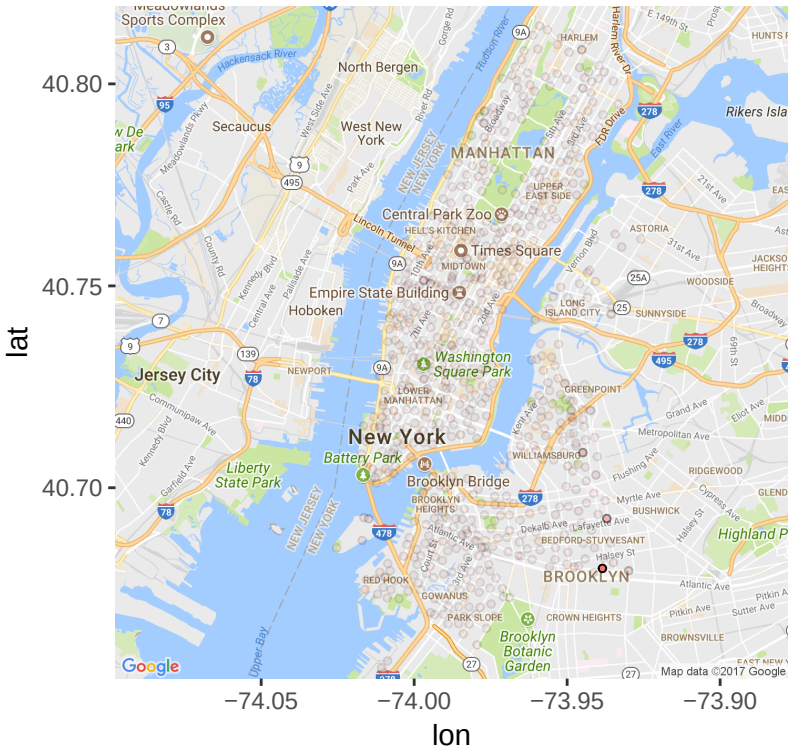
Count visualization Map

```
ggmap(mymap) +  geom_point(data = start_sums,
                           aes(x = start.station.longitude,
                               y = start.station.latitude,
                               fill = "red", alpha = start.count),
                           size = 1, shape = 21) +
  guides(fill=FALSE, alpha=FALSE, size=FALSE)
```



Avg. Trip Time Map

```
ggmap(mymap) +  geom_point(data = start_sums,
                           aes(x = start.station.longitude,
                               y = start.station.latitude,
                               fill = "red", alpha = avg.time.out),
                           size = 1, shape = 21) +
  guides(fill=FALSE, alpha=FALSE, size=FALSE)
```




```
start_sums %>%
  arrange(desc(start.count)) %>%
  select(-start.station.latitude, -start.station.longitude) %>%
  head()
```

```
## # A tibble: 6 x 3
##   start.station.id start.count avg.time.out
##              <int>       <int>        <dbl>
## 1              519        5367        838.2
## 2              497        4057        749.4
## 3              435        3948        658.4
## 4              426        3691       1221.8
## 5              402        3482        743.2
## 6              293        3461        720.4
```

```
start_sums %>%
  arrange(desc(avg.time.out)) %>%
  select(-start.station.latitude, -start.station.longitude) %>%
  head()
```

```
## # A tibble: 6 x 3
##   start.station.id start.count avg.time.out
##              <int>       <int>        <dbl>
## 1             3044          29        61691
## 2             3058          91        28080
## 3             3076         125        22146
## 4             3042         118        12161
## 5             3518           7        11179
## 6             3342          26         7160
```

```
end_sums <- randomsample %>%
  group_by(end.station.id) %>%
```

```
  summarize(end.station.longitude = mean(end.station.longitude),
            end.station.latitude = mean(end.station.latitude),
            total.time.in = sum(tripduration),
            end.count = n()) %>%
  mutate(avg.time.in = total.time.in/end.count) %>%
  select(-total.time.in) %>%
  ungroup()
```

```
colnames(start_sums)[1]<- c("id")
colnames(end_sums)[1] <- c("id")
joined_data<-left_join(start_sums,end_sums,by="id")
joined_data <- joined_data %>%
mutate(difference = start.count - end.count,
       station.latitude =
         (start.station.latitude+ end.station.latitude)/2,
       station.longitude =
         (start.station.longitude + end.station.longitude)/2) %>%
  select(-start.station.latitude,
         -start.station.longitude,
         -end.station.longitude,
         -end.station.latitude)
```

```
biggest_differences <- joined_data %>%
  arrange(desc(difference)) %>%
head(10)
biggest_differences %>%
    select(station.latitude, station.longitude)
```

```
## # A tibble: 10 x 2
##    station.latitude station.longitude
##               <dbl>             <dbl>
##  1            40.75            -73.98
##  2            40.76            -73.99
##  3            40.75            -73.98
##  4            40.75            -74.00
##  5            40.75            -73.99
##  6            40.77            -73.98
##  7            40.76            -73.99
##  8            40.76            -73.97
##  9            40.75            -73.99
## 10            40.76            -73.99
```

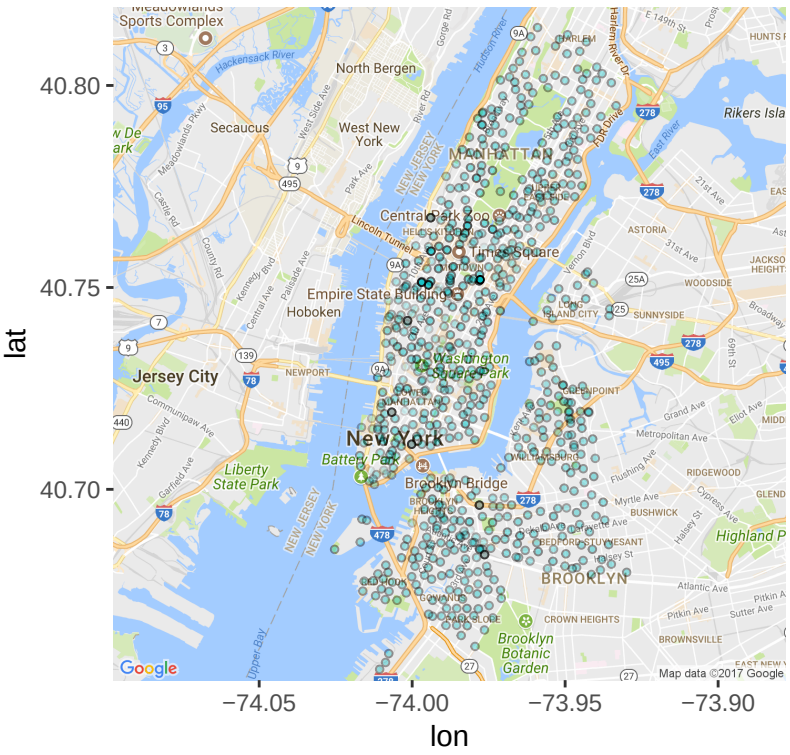
```
smallest_differences <- joined_data %>%
  arrange(difference) %>%
head(10)
smallest_differences %>%
  select(station.latitude, station.longitude)
```

```
## # A tibble: 10 x 2
##    station.latitude station.longitude
##               <dbl>             <dbl>
##  1            40.73            -73.99
##  2            40.76            -73.98
##  3            40.69            -73.98
##  4            40.70            -74.01
```

```
##  5            40.75            -73.99
##  6            40.76            -74.00
##  7            40.73            -73.98
##  8            40.73            -74.00
##  9            40.74            -73.99
## 10            40.72            -74.00
```

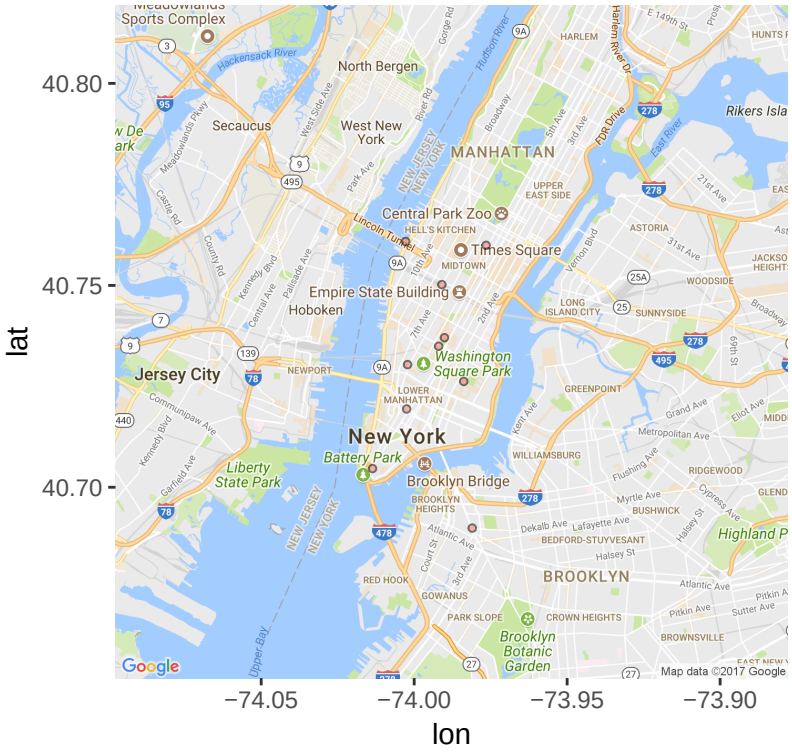
Difference visualization Map

```
ggmap(mymap) +  geom_point(data = joined_data,
                           aes(x = station.longitude,
                               y = station.latitude,
                               fill = "red", alpha = difference),
                           size = 1, shape = 21) +
  guides(fill=FALSE, alpha=FALSE, size=FALSE) +
  geom_point(data=biggest_differences,
             aes(x = station.longitude,
                 y = station.latitude,
                 fill = "red",
                 alpha = 1.0),
             size = 1, shape = 21) +
    geom_point(data=smallest_differences,
             aes(x = station.longitude,
                 y = station.latitude,
                 fill = "blue",
                 alpha = 1.0),
             size = 1, shape = 21)
```

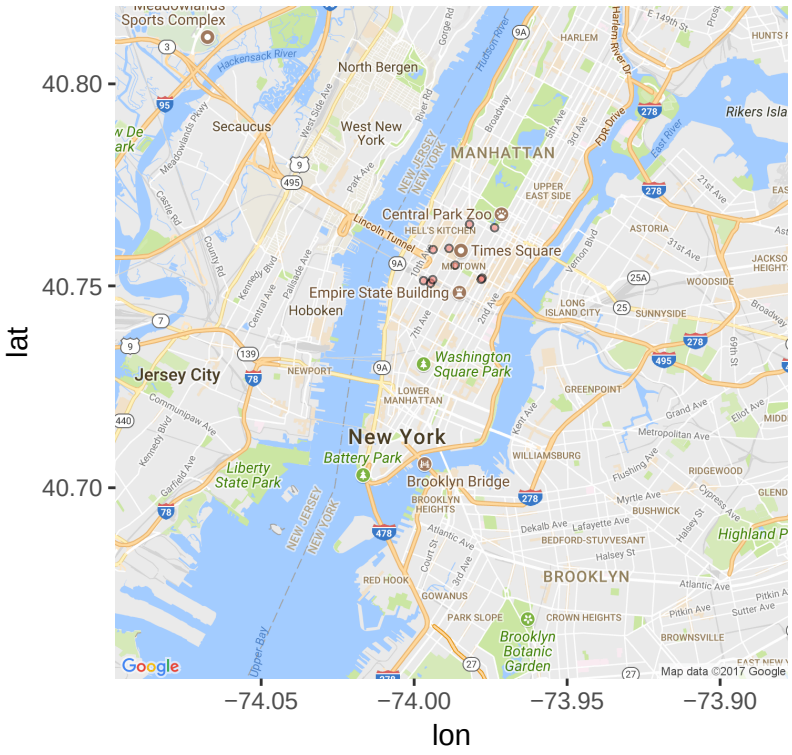



```
ggmap(mymap) +
    geom_point(data=smallest_differences,
             aes(x = station.longitude,
```

```
                y = station.latitude,
                fill = "blue",
                alpha = 1.0),
            size = 1, shape = 21) +
  guides(fill=FALSE, alpha=FALSE, size=FALSE)
```




```
ggmap(mymap) +
  geom_point(data=biggest_differences,
             aes(x = station.longitude,
                 y = station.latitude,
                 fill = "red",
                 alpha = 1.0),
             size = 1, shape = 21) +
  guides(fill=FALSE, alpha=FALSE, size=FALSE)
```


lat
40.80
40.75
40.70
−74.05
−74.00
−73.95
−73.90
lon
Sports Complex
Hackensack River
North Bergen
Secaucus
West New York
MANHATTAN
HARLEM
Rikers Isla
Central Park Zoo
HELL'S KITCHEN
Times Square
MIDTOWN
Empire State Building
Hoboken
Jersey City
NEWPORT
Washington Square Park
LOWER MANHATTAN
New York
Battery Park
Liberty State Park
Brooklyn Bridge
WILLIAMSBURG
GREENPOINT
BROOKLYN HEIGHTS
BEDFORD-STUYVESANT
BROOKLYN
RED HOOK
GOWANUS
PARK SLOPE
CROWN HEIGHTS
Brooklyn Botanic Garden
BUSHWICK
RIDGEWOOD
ASTORIA
LONG ISLAND CITY
SUNNYSIDE
WOODSIDE
Upper Bay
Google
Map data ©2017 Google