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, start.station.name,
           start.station.latitude, start.station.longitude) %>%
  summarize(total.time.out = sum(tripduration),start.count = n())

start_sums <- start_sums %>%
  mutate(avg.time.out = total.time.out/start.count) %>%
  select(-total.time.out)

start_sums <- start_sums %>% ungroup
```

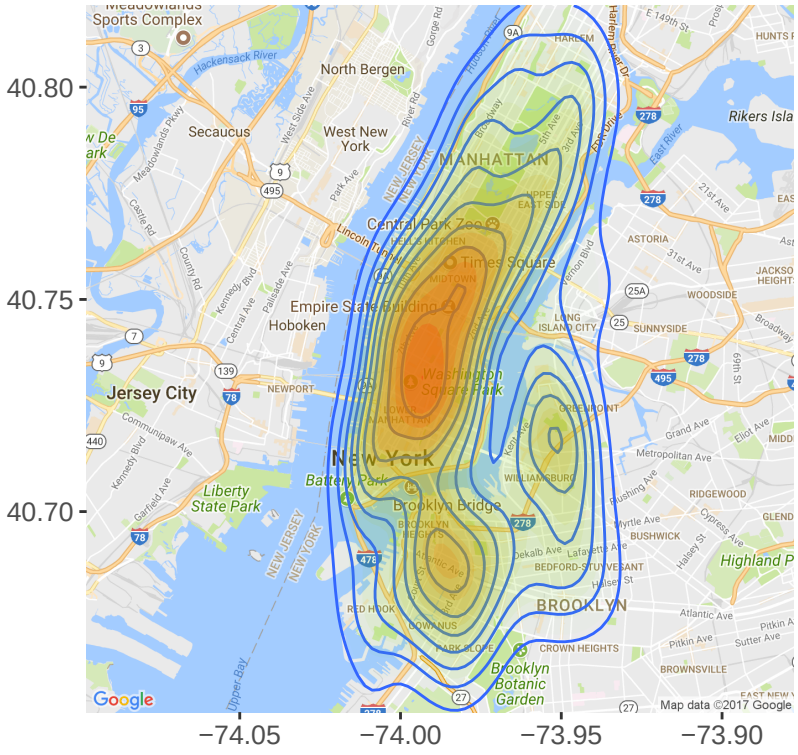
```
end_sums <- randomsample %>%
  group_by(end.station.id, end.station.name,
           end.station.latitude, end.station.longitude) %>%
  summarize(total.time.in = sum(tripduration),end.count = n())

end_sums <- end_sums %>%
  mutate(avg.time.in = total.time.in/end.count) %>%
  select(-total.time.in)

end_sums <- end_sums %>% ungroup
```

```
head(start_sums)
```

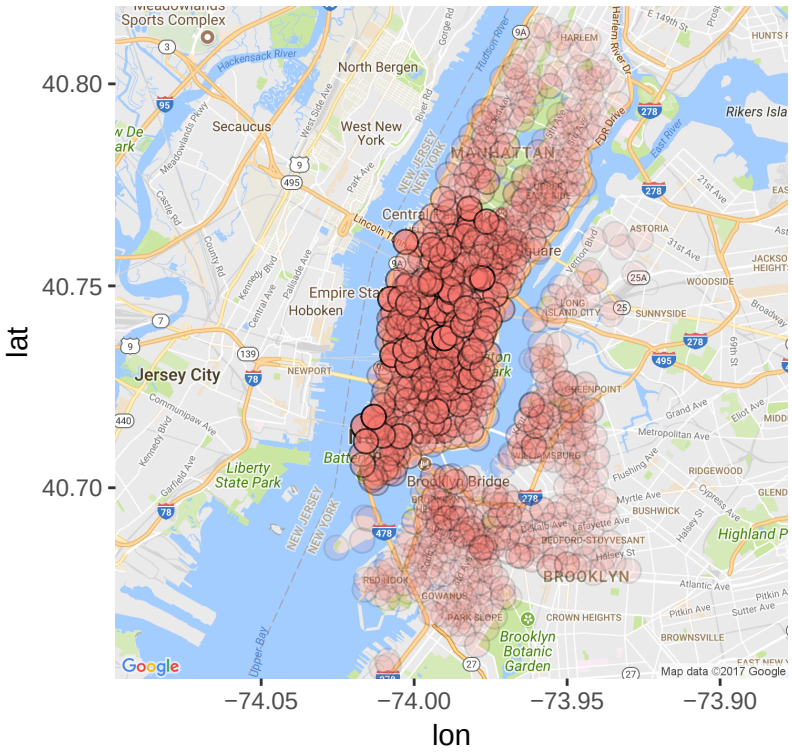
```
## # A tibble: 6 x 6
##   start.station.id           start.station.name start.station.latitude
##              <int>                        <chr>                  <dbl>
## 1               72             W 52 St & 11 Ave                  40.77
## 2               79     Franklin St & W Broadway                  40.72
## 3               82       St James Pl & Pearl St                  40.71
## 4               83 Atlantic Ave & Fort Greene Pl                  40.68
## 5              116              W 17 St & 8 Ave                  40.74
## 6              119     Park Ave & St Edwards St                  40.70
## # ... with 3 more variables: start.station.longitude <dbl>,
## #   start.count <int>, avg.time.out <dbl>
```

```
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)
```

```
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))
```




```
ggmap(mymap) +  geom_point(data = start_sums,
                        aes(x = start.station.longitude,
                            y = start.station.latitude,
                            fill = "red", alpha = start.count),
                        size = 4, 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 4
##   start.station.id    start.station.name start.count avg.time.out
##              <int>                 <chr>       <int>        <dbl>
## 1              519 Pershing Square North        4170        849.0
## 2              497    E 17 St & Broadway        4057        749.4
## 3              435       W 21 St & 6 Ave        3948        658.4
## 4              426 West St & Chambers St        3691       1221.8
## 5              402    Broadway & E 22 St        3482        743.2
## 6              285    Broadway & E 14 St        3308        715.3
```

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

‘‘‘