airbnb_project

Matthew Coleman

2/28/2020

```r
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------------------------- tidyve
```

```
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.2     v dplyr   0.8.1
## v tidyr   0.8.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ---------------------------------------------------------------------------- tidyverse_co
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

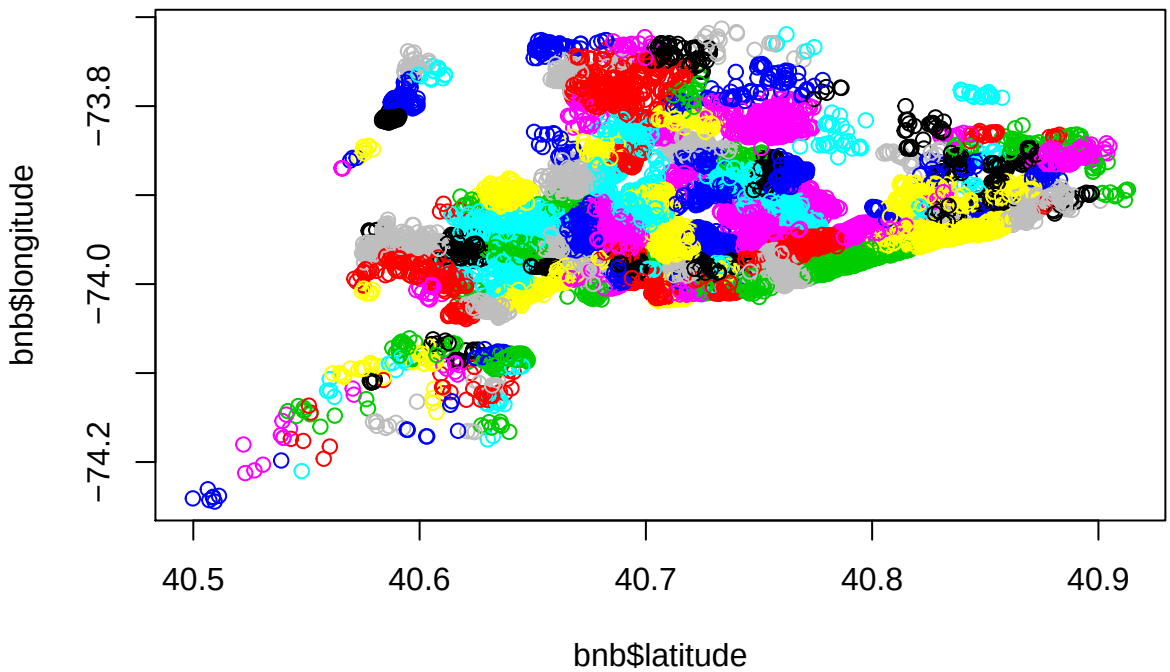
```r
library(dplyr)

bnb <- read.csv('AB_NYC_2019.csv')

bnb[is.na(bnb$room_type),]
```

```
##  [1] id                              name
##  [3] host_id                         host_name
##  [5] neighbourhood_group             neighbourhood
##  [7] latitude                        longitude
##  [9] room_type                       price
## [11] minimum_nights                  number_of_reviews
## [13] last_review                     reviews_per_month
## [15] calculated_host_listings_count availability_365
## <0 rows> (or 0-length row.names)
```

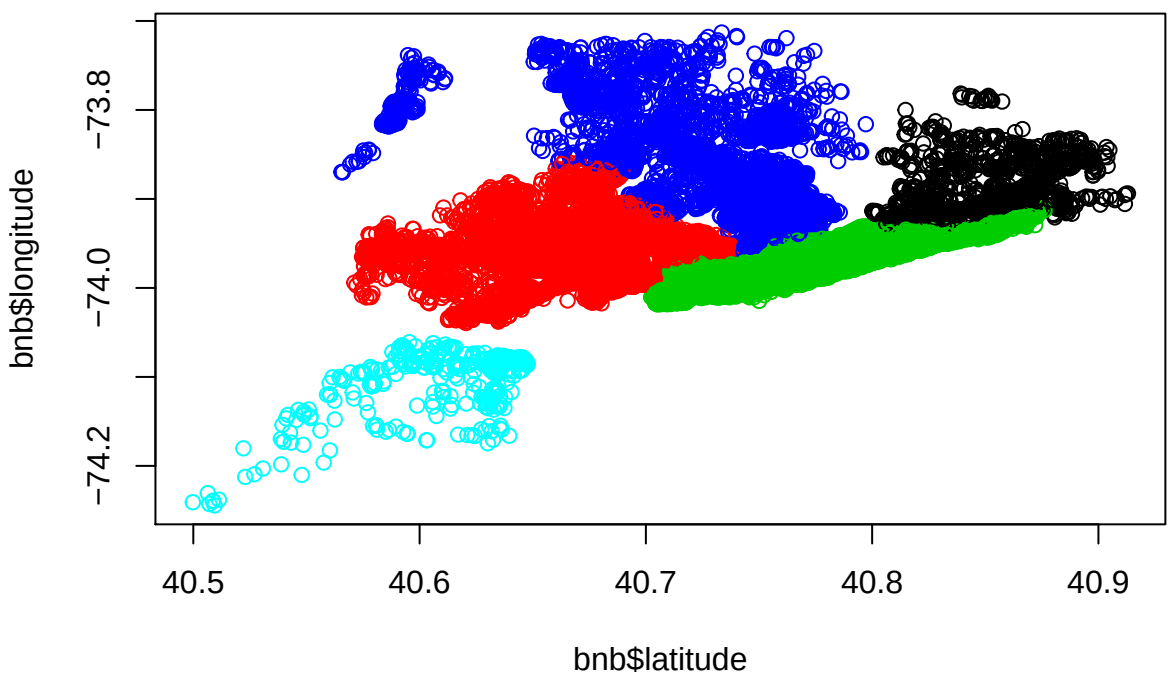
```r
n_distinct(bnb$neighbourhood_group)
```

```
## [1] 5
```

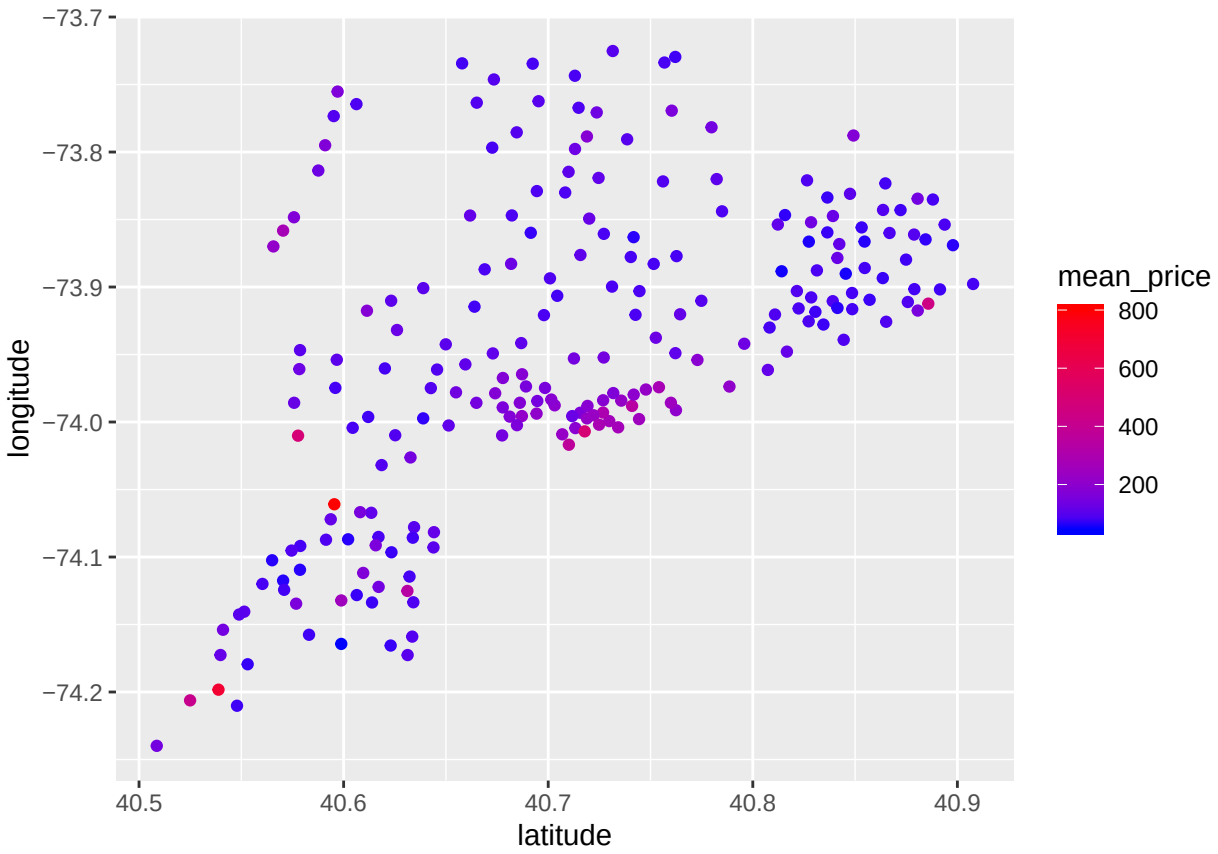
```r
plot(bnb$latitude, bnb$longitude, col = bnb$neighbourhood)
```




```
plot(bnb$latitude, bnb$longitude, col = bnb$neighbourhood_group)
```




```
mean_price <- bnb %>% group_by(neighbourhood) %>% summarise(mean_price = mean(price),
                                                            latitude = median(latitude), longitude =
                                                            median(longitude))

#plot(mean_price$latitude, mean_price$longitude, col = mean_price$mean_price)

ggplot(data = mean_price, aes(x = latitude, y = longitude, color = mean_price)) +
  geom_point() + scale_color_gradient(low="blue", high="red")
```


−73.7
−73.8
−73.9
−74.0
−74.1
−74.2
longitude
40.5
40.6
40.7
40.8
40.9
latitude
mean_price
800
600
400
200