Lab7

Suixin Jiang

10/27/2019

Exercise 1

```
data("flights")
head(flights)
```

```
## # A tibble: 6 x 19
##    year month   day dep_time sched_dep_time dep_delay arr_time
##   <int> <int> <int>    <int>          <int>     <dbl>    <int>
## 1  2013     1     1      517            515         2      830
## 2  2013     1     1      533            529         4      850
## 3  2013     1     1      542            540         2      923
## 4  2013     1     1      544            545        -1     1004
## 5  2013     1     1      554            600        -6      812
## 6  2013     1     1      554            558        -4      740
## # ... with 12 more variables: sched_arr_time <int>, arr_delay <dbl>,
## #   carrier <chr>, flight <int>, tailnum <chr>, origin <chr>, dest <chr>,
## #   air_time <dbl>, distance <dbl>, hour <dbl>, minute <dbl>,
## #   time_hour <dttm>
```

Exercise 2

The location of NYC is (40.70317N, 73.94007W)

```
data("airports")
head(airports)
```

```
## # A tibble: 6 x 8
##   faa   name                       lat   lon   alt    tz dst   tzone
##   <chr> <chr>                    <dbl> <dbl> <dbl> <dbl> <chr> <chr>
## 1 04G   Lansdowne Airport         41.1 -80.6  1044    -5 A     America/New~
## 2 06A   Moton Field Municipal A~  32.5 -85.7   264    -6 A     America/Chi~
## 3 06C   Schaumburg Regional       42.0 -88.1   801    -6 A     America/Chi~
## 4 06N   Randall Airport           41.4 -74.4   523    -5 A     America/New~
## 5 09J   Jekyll Island Airport     31.1 -81.4    11    -5 A     America/New~
## 6 0A9   Elizabethton Municipal ~  36.4 -82.2  1593    -5 A     America/New~
```

```
flights %>%
  select(origin) %>%
  distinct() %>%
  left_join(airports, by=c('origin'='faa')) %>%
  summarize(ave_lat=mean(lat), ave_lon=mean(lon)) ->
  nyc_location
nyc_location
```

```
## # A tibble: 1 x 2
##   ave_lat ave_lon
##     <dbl>   <dbl>
```

```
## 1    40.7   -73.9
```

Exercise 3

For distm(x, y, fun=distGeo) function, arguments x has to be with longitude first and then latitude.

```
lon <- c(-80.6, -85.7, -88.1)
lat <- c(41.1, 32.5, 42.0)
nyclon <- -73.9
nyclat <- 40.8
geo_dist = function(lon, lat, nyclon, nyclat) {
  distm(cbind(lon,lat),cbind(nyclon,nyclat),fun = distCosine)
}
geo_dist(lon=lon, lat=lat, nyclon=nyclon, nyclat=nyclat)
```

```
##           [,1]
## [1,]  564169.1
## [2,] 1399198.9
## [3,] 1191834.4
```

Exercise 4

The plot shows that the more distance from NYC the less average arrival delay time. Low flight density reduces unnecessary air traffic control.

```
airports %>%
  mutate(dist_nyc=geo_dist(lon=lon,lat=lat,nyclon=nyc_location$ave_lon,nyclat=nyc_location$ave_lat)) ->
  airports
airports
```

```
## # A tibble: 1,458 x 9
##    faa   name            lat    lon   alt    tz dst   tzone    dist_nyc[,1]
##    <chr> <chr>         <dbl>  <dbl> <dbl> <dbl> <chr> <chr>           <dbl>
##  1 04G   Lansdowne Ai~  41.1  -80.6  1044    -5 A     America~      563749.
##  2 06A   Moton Field ~  32.5  -85.7   264    -6 A     America~     1391697.
##  3 06C   Schaumburg R~  42.0  -88.1   801    -6 A     America~     1190702.
##  4 06N   Randall Airp~  41.4  -74.4   523    -5 A     America~       89537.
##  5 09J   Jekyll Islan~  31.1  -81.4    11    -5 A     America~     1265667.
##  6 0A9   Elizabethton~  36.4  -82.2  1593    -5 A     America~      863399.
##  7 0G6   Williams Cou~  41.5  -84.5   730    -5 A     America~      890110.
##  8 0G7   Finger Lakes~  42.9  -76.8   492    -5 A     America~      338361.
##  9 0P2   Shoestring A~  39.8  -76.6  1000    -5 U     America~      251238.
## 10 0S9   Jefferson Co~  48.1 -123.    108    -8 A     America~     3905014.
## # ... with 1,448 more rows
```

```
flights %>%
  group_by(dest) %>%
  summarize(ave_arr_delay=mean(arr_delay, na.rm = T)) %>%
  left_join(airports, by=c('dest'='faa')) %>%
  ggplot(mapping=aes(x=dist_nyc, y=ave_arr_delay)) +
  geom_point() +
  geom_smooth(method = lm, se = F) +
```

```
  theme_bw() +
  xlab('Distance from NYC') +
  scale_y_continuous('Average Arrival Dealy')
```

```
## Warning: Removed 5 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```