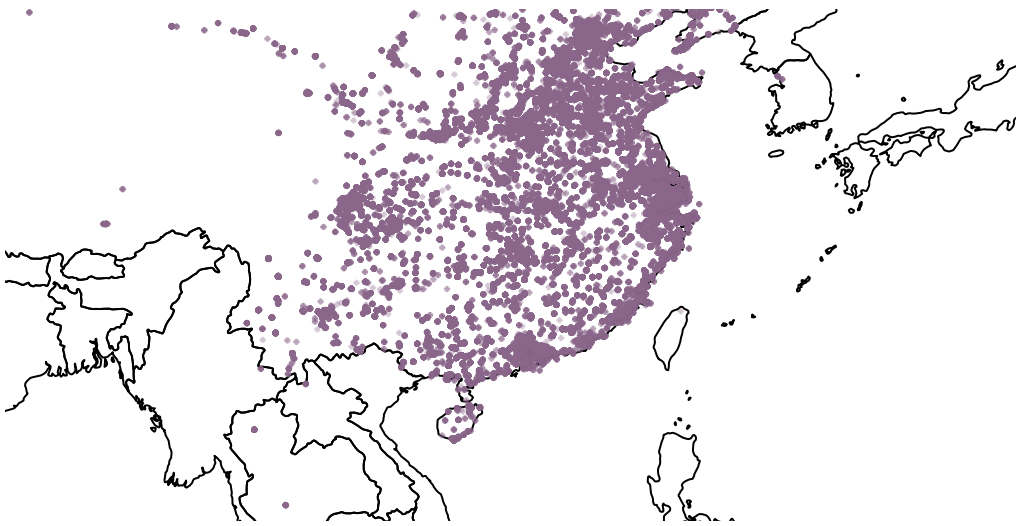



```
# MAY 8
plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1, main = "mobile activity: may 8")
points(may_8$longitude, may_8$latitude, col = alpha("plum4", 0.2), cex = 0.3, pch = 10)
```

mobile activity: may 8

NEED GOOGLE API KEY FOR THIS MAP

```
register_google(key = 'AIzaSyAj98ZG3Tt8xFIdx78HRi3MMb1LRB_ZYfQ', "standard", client = "uc-berkeley-urba

has_google_key()
```

```
china <- ggmap(get_googlemap(center = c(lon = 112.363625, lat = 32),
                    zoom = 5, scale = 2,
                    maptype ='roadmap',
                    color = 'color'))
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=32,112.363625&zoom=5&size=640x640&scal
```

```
spatial_provinces <- readOGR("Archive/provinces")
```

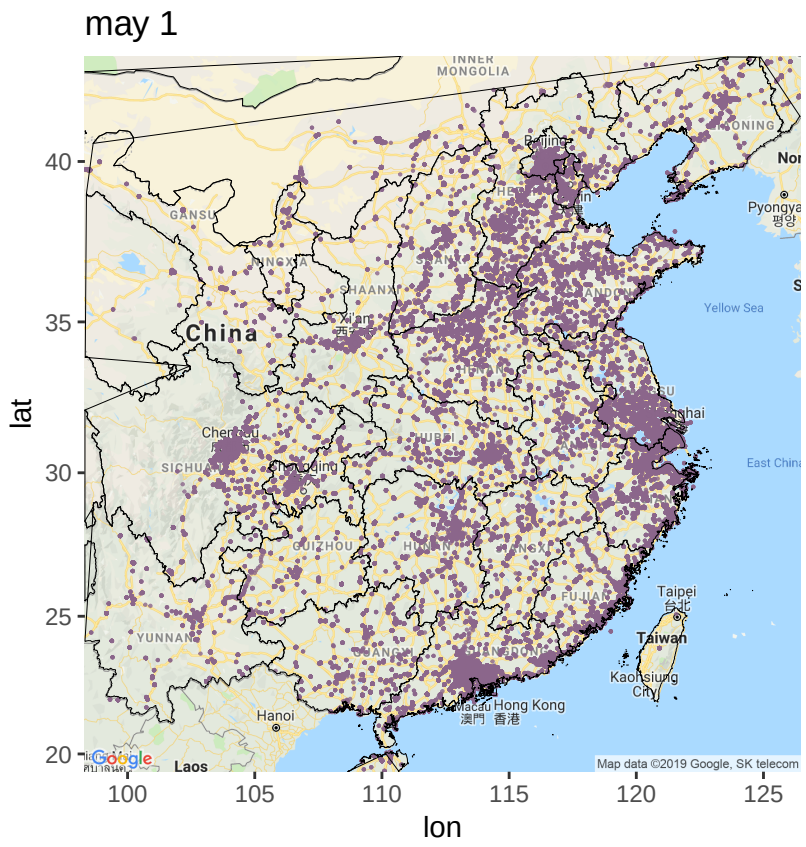
```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/pet/Documents/caL/2019/cyplan101/projects/assignment3/Archive/provinces", layer: "nyu
## with 925 features
## It has 6 fields
```

```
spatial_provinces <- spTransform(spatial_provinces, CRS("+proj=longlat +datum=WGS84"))

provinces <- fortify(spatial_provinces)
```

```
## Regions defined for each Polygons
```

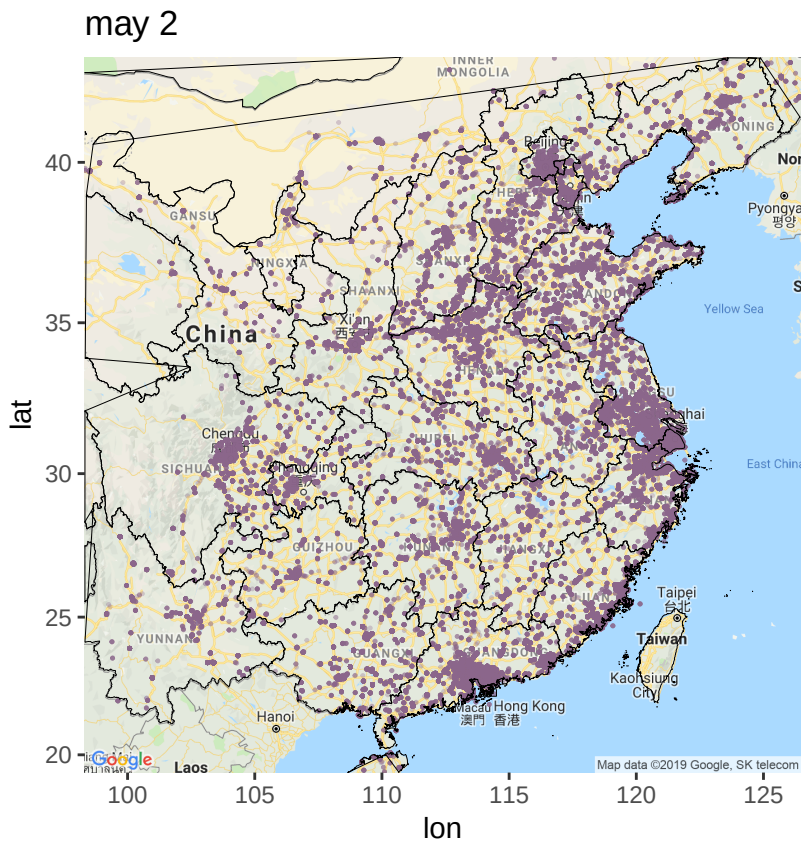
```
five_prov <- provinces[provinces$id %in% c(11, 31, 52, 45, 21), ]

china + geom_point(aes(x = longitude, y = latitude), data = may_1, size = 0.1, col = "plum4", alpha = 0
```

```
## Warning: Removed 133581 rows containing missing values (geom_point).
```



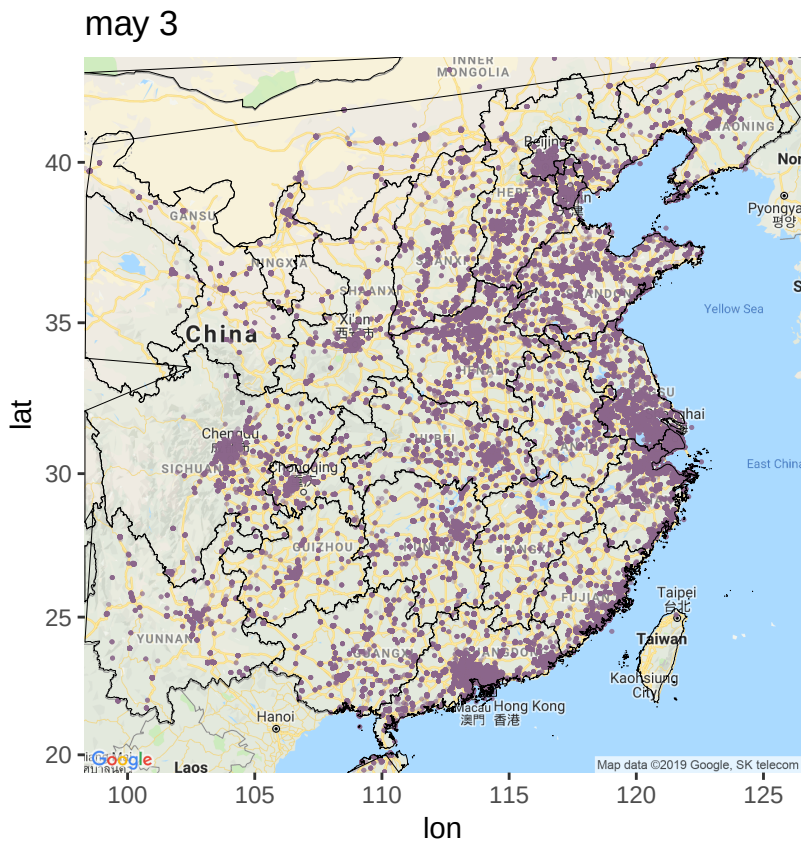

```
china + geom_point(aes(x = longitude, y = latitude), data = may_2, size = 0.2, col = "plum4", alpha = 0
```

```
## Warning: Removed 143134 rows containing missing values (geom_point).
```



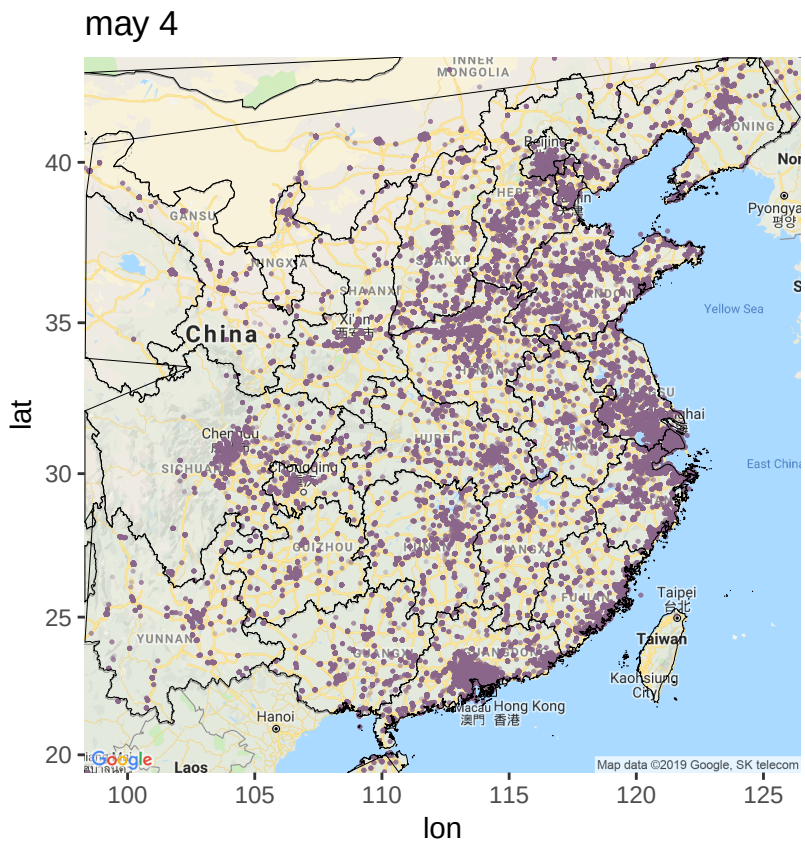

```
china + geom_point(aes(x = longitude, y = latitude), data = may_3, size = 0.2, col = "plum4", alpha = 0
```

```
## Warning: Removed 139702 rows containing missing values (geom_point).
```



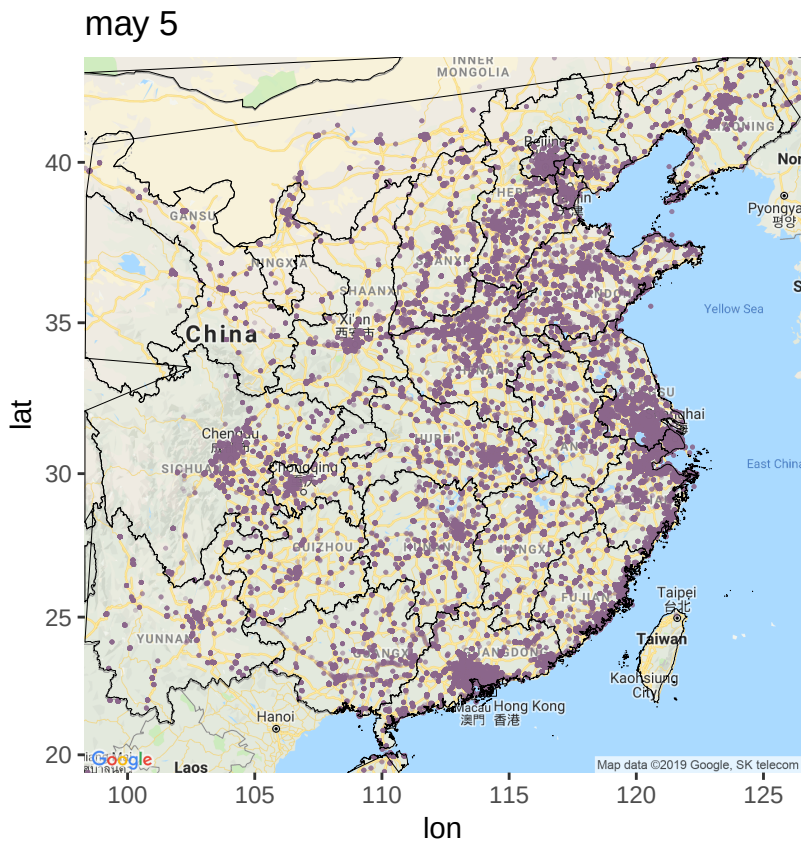

```
china + geom_point(aes(x = longitude, y = latitude), data = may_4, size = 0.2, col = "plum4", alpha = 0
```

```
## Warning: Removed 145330 rows containing missing values (geom_point).
```



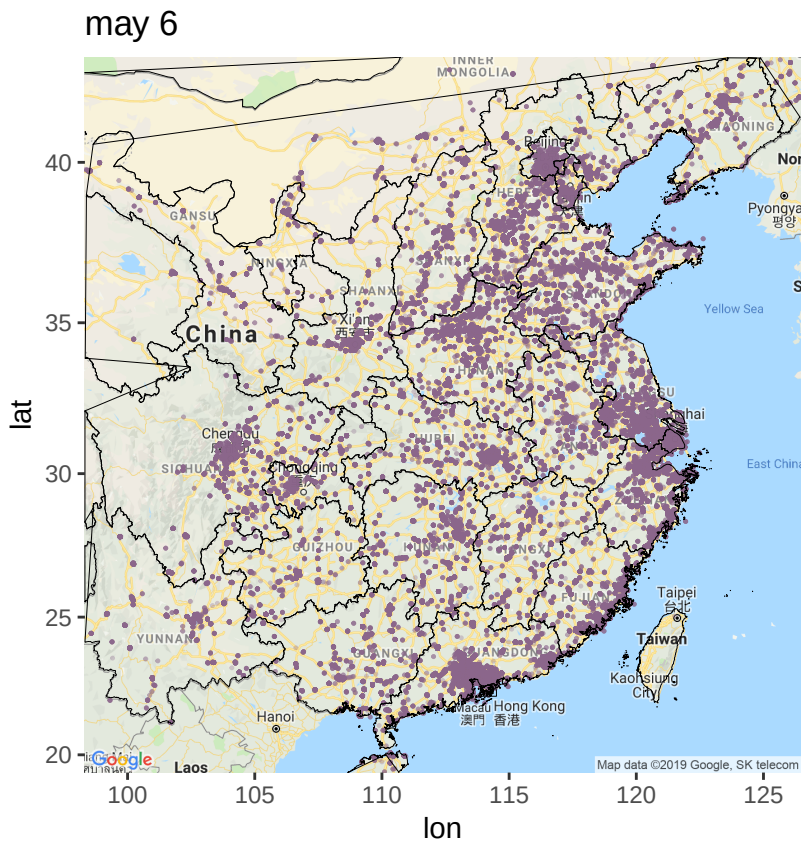

```
china + geom_point(aes(x = longitude, y = latitude), data = may_5, size = 0.2, col = "plum4", alpha = 0
```

```
## Warning: Removed 147263 rows containing missing values (geom_point).
```



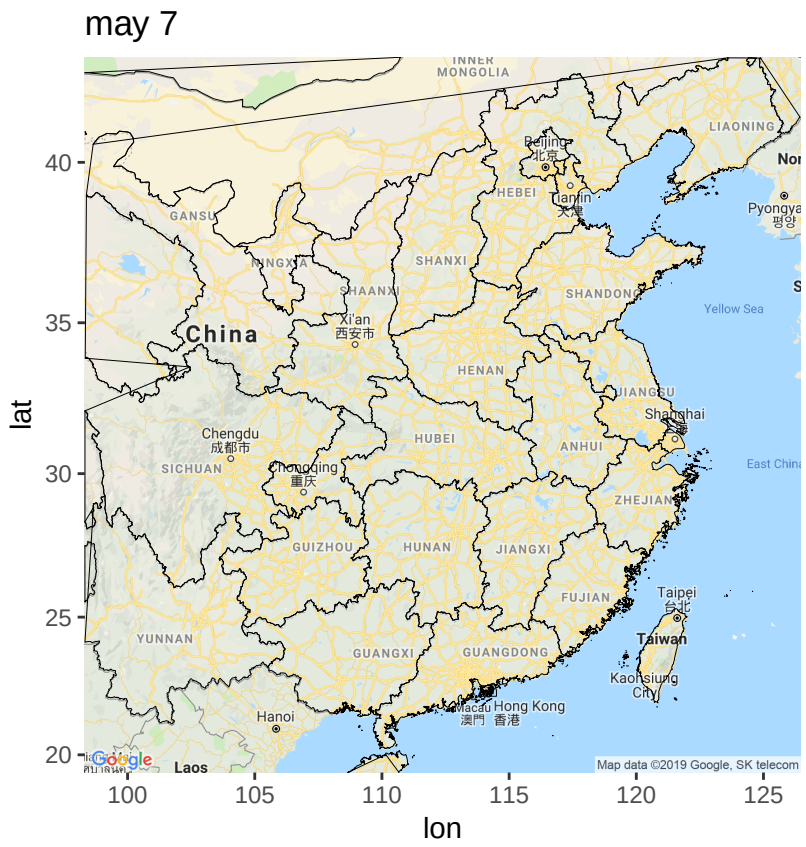

```
china + geom_point(aes(x = longitude, y = latitude), data = may_6, size = 0.2, col = "plum4", alpha = 0
```

```
## Warning: Removed 137721 rows containing missing values (geom_point).
```




```
china + geom_point(aes(x = longitude, y = latitude), data = may_7, size = 0.2, col = "plum4", alpha = 0
```

```
## Warning: Removed 3032372 rows containing missing values (geom_point).
```




```
may_1


p_may1 <- SpatialPointsDataFrame(may_1[,c(3,4)], may_1)
p_may2 <- SpatialPointsDataFrame(may_2[,c(3,4)], may_2)
p_may3 <- SpatialPointsDataFrame(may_3[,c(3,4)], may_3)
p_may4 <- SpatialPointsDataFrame(may_4[,c(3,4)], may_4)
p_may5 <- SpatialPointsDataFrame(may_5[,c(3,4)], may_5)
p_may6 <- SpatialPointsDataFrame(may_6[,c(3,4)], may_6)
p_may7 <- SpatialPointsDataFrame(may_7[,c(3,4)], may_7)


proj4string(spatial_provinces) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")

proj4string(p_may1) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
proj4string(p_may2) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
proj4string(p_may3) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
proj4string(p_may4) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
proj4string(p_may5) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
proj4string(p_may6) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")
proj4string(p_may7) <- CRS("+proj=longlat +ellps=WGS84 +datum=WGS84")

res_51 <- over(p_may1, spatial_provinces)
res_52 <- over(p_may2, spatial_provinces)
res_53 <- over(p_may3, spatial_provinces)
res_54 <- over(p_may4, spatial_provinces)
```

```
res_55 <- over(p_may5, spatial_provinces)
res_56 <- over(p_may6, spatial_provinces)
res_57 <- over(p_may7, spatial_provinces)


sort(as.integer(names(table(res_51$prov_id))))


count1 <- table(res_51$prov_id)[names(table(res_51$prov_id)) %in% c(11,31,52,45,21)]
count2 <- table(res_52$prov_id)[names(table(res_52$prov_id)) %in% c(11,31,52,45,21)]
count3 <- table(res_53$prov_id)[names(table(res_53$prov_id)) %in% c(11,31,52,45,21)]
count4 <- table(res_54$prov_id)[names(table(res_54$prov_id)) %in% c(11,31,52,45,21)]
count5 <- table(res_55$prov_id)[names(table(res_55$prov_id)) %in% c(11,31,52,45,21)]
count6 <- table(res_56$prov_id)[names(table(res_56$prov_id)) %in% c(11,31,52,45,21)]
count7 <- table(res_57$prov_id)[names(table(res_57$prov_id)) %in% c(11,31,52,45,21)]

activity_counts <- data.frame(
  may_1 = count1,
  may_2 = count2,
  may_3 = count3,
  may_4 = count4,
  may_5 = count5,
  may_6 = count6,
  may_7 = count7
)


activity_counts <- activity_counts[,c(2, 4, 6, 8, 10, 12, 14)]


#write.csv(activity_counts, "activity_counts.csv")

# beijing : 11, 31, 52, 45, 21


# beijing, shenyang, chengdu, guangzhou


#beijing, shenyang, shanghai, guangzhou, chengdu
```

% total checkins in each city

$ total checkins of all days in each city

CLUSTERING

```
knn_events <- kmeans(events[,c(4,5)], 20, nstart = 20)

kmeans_51 <- kmeans(may_1[,c(3,4)], 20, nstart = 20)
```

```
kmeans_52 <- kmeans(may_2[,c(3,4)], 20, nstart = 20)
kmeans_53 <- kmeans(may_3[,c(3,4)], 20, nstart = 20)
kmeans_54 <- kmeans(may_4[,c(3,4)], 20, nstart = 20)
kmeans_55 <- kmeans(may_5[,c(3,4)], 20, nstart = 20)
kmeans_56 <- kmeans(may_6[,c(3,4)], 20, nstart = 20)
kmeans_57 <- kmeans(may_7[,c(3,4)], 20, nstart = 20)
```

plots

```
newmap <- getMap(resolution = "low")
plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1)
points(events$longitude, events$latitude, col = "rosybrown2", cex = 0.05)
points(knn_events$centers[,1], knn_events$centers[,2], col = "red", pch = 19)
#summary(events$longitude)
#summary(events$latitude)

centers <- data.frame(
  events_long = knn_events$centers[,1],
  events_lat = knn_events$centers[,2],
  long_51 = kmeans_51$centers[,1],
  lat_51 = kmeans_51$centers[,2],
  long_52 = kmeans_52$centers[,1],
  lat_52 = kmeans_52$centers[,2],
  long_53 = kmeans_53$centers[,1],
  lat_53 = kmeans_53$centers[,2],
  long_54 = kmeans_54$centers[,1],
  lat_54 = kmeans_54$centers[,2],
  long_55 = kmeans_55$centers[,1],
  lat_55 = kmeans_55$centers[,2],
  long_56 = kmeans_56$centers[,1],
  lat_56 = kmeans_56$centers[,2],
  long_57 = kmeans_57$centers[,1],
  lat_57 = kmeans_57$centers[,2]
)

#write.csv(centers, "centers.csv")
```

```
plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1, main = "may 1 clusters")
points(may_1$longitude, may_1$latitude, col = "rosybrown2", cex = 0.05)
points(kmeans_51$centers[,1], kmeans_51$centers[,2], col = "red", pch = 19)

plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1, main = "may 2 clusters")
points(may_2$longitude, may_2$latitude, col = "rosybrown2", cex = 0.05)
points(kmeans_52$centers[,1], kmeans_52$centers[,2], col = "red", pch = 19)

plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1, main = "may 3 clusters")
points(may_3$longitude, may_3$latitude, col = "rosybrown2", cex = 0.05)
points(kmeans_53$centers[,1], kmeans_53$centers[,2], col = "red", pch = 19)

plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1, main = "may 4 clusters")
points(may_4$longitude, may_4$latitude, col = "rosybrown2", cex = 0.05)
points(kmeans_54$centers[,1], kmeans_54$centers[,2], col = "red", pch = 19)
```

```r
plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1, main = "may 5 clusters")
points(may_5$longitude, may_5$latitude, col = "rosybrown2", cex = 0.05)
points(kmeans_55$centers[,1], kmeans_55$centers[,2], col = "red", pch = 19)

plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1, main = "may 6 clusters")
points(may_6$longitude, may_6$latitude, col = "rosybrown2", cex = 0.05)
points(kmeans_56$centers[,1], kmeans_56$centers[,2], col = "red", pch = 19)

plot(newmap, xlim = c(105, 120), ylim = c(15, 40), asp = 1, main = "may 7 clusters")
points(may_7$longitude, may_7$latitude, col = "rosybrown2", cex = 0.05)
points(kmeans_57$centers[,1], kmeans_57$centers[,2], col = "red", pch = 19)
```