Project 2. Spatial Exploration of Tweets

Abstract

This project practiced the sp and sf object creation, in and output of spatial object. In addition, this project practiced the mapping distribution of month, weekofday, and hours. This project also analyzed the distribution of selecting top three users and conducted spatial temporal analysis.

1. Data Preparation – sp and sf object creation

This initial section of the project asked the users to read the files, create sp, and sf object. After the initial step, the users are asked to create the bounding box for the sp and sf object. During this step, I added extra space for the latitude and longitude. Referring to the comment from the in-progress report the extended latitude and longitude have been modified to 0.01. The following figure represents the code of adding extra latitude and longitude.

```
#creat the bounding box
#find the minimum and maximum latitude and longitude coordinates of tweets
xmin <- min(pu2014$longitude)
ymin <- min(pu2014$latitude)
xmax <- max(pu2014$longitude)
ymax <- max(pu2014$latitude)

#made a bit extra space for the coordinates, 0.01 for latitude and longitude
extra_latlong <- 0.01

#expand the latitude and longitude for 'sp' object
#calculate new bounding box coordinates
new_xmin_sp <- xmin - extra_latlong
new_ymin_sp <- ymin - extra_latlong
new_xmax_sp <- xmax + extra_latlong
new_ymax_sp <- ymax + extra_latlong
```

2. Spatial object I/O and conversion

This part of the project requires write, convert of the files, one thing needs to be careful is that the file needs to be able to overwrite in order to rerun the current program. In addition to that, I also convert the file to KM, Geojson, and Geopackage. (screenshot in reference).

```
#write the 'sp' object to an ESRI Shapefile using rgdal
rgdal::writeOGR(obj = sp, dsn = sp_out_path, layer = layer_name_sp, driver = 'ESRI Shapefile', overwrite_layer = TRUE)

#write the 'sf' object to an ESRI Shapefile using sf
sf::st_write(sf, dsn = sf_out_path, layer = layer_name_sf, driver = 'ESRI Shapefile', append = FALSE)
```

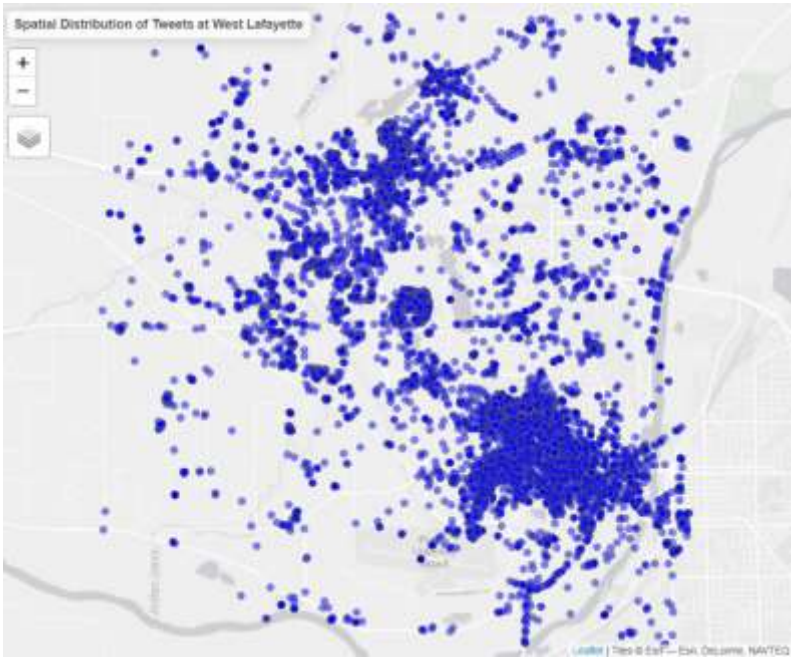
3. Spatial object's attribute manipulation (use the sf file from Task 1.1)

This part of the project asked users to convert the epoch time today month year, day of time and the local time. After that, those converted time will be added to sf, after cleaning the data, I have selected the following subset into the new sf file for the future analysis. the graph below shows the columns I selected for the future analysis.

```
Bounding box:  xmin: -86.99292 ymin: 40.42621 xmax: -86.9006 ymax: 40.47966
Geodetic CRS:  WGS 84
    user_id                  geometry      epoch            datetime       date dayOfweek   localTime      lat
1 174220305 POINT (-86.94425 40.47112) 1388552464 2014-01-01 00:01:04 01-01-2014         3 12:01:04 AM 40.47112
2  99818152 POINT (-86.94266 40.44576) 1388552467 2014-01-01 00:01:07 01-01-2014         3 12:01:07 AM 40.44576
3  30137074 POINT (-86.93918 40.47966) 1388552533 2014-01-01 00:02:13 01-01-2014         3 12:02:13 AM 40.47966
4 458797537  POINT (-86.99292 40.4582) 1388552645 2014-01-01 00:04:05 01-01-2014         3 12:04:05 AM 40.45820
5  26849093  POINT (-86.9006 40.42621) 1388552648 2014-01-01 00:04:08 01-01-2014         3 12:04:08 AM 40.42621
       lon
1 -86.94425
2 -86.94266
3 -86.93918
4 -86.99292
5 -86.90060
```

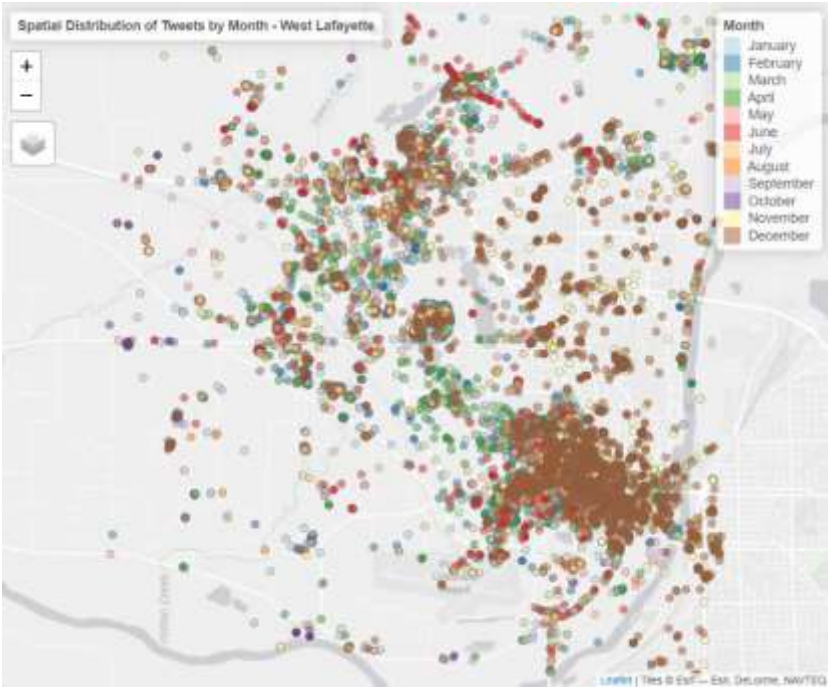
4. Mapping the overall tweets distribution

This section asked users to show the overall distribution of tweet locations using either Open Street Map (OSM) or the Google Map. I have tried to use the tmap function to analyze the overall distribution of the tweet. The following map shows the overall distribution of the tweet near West

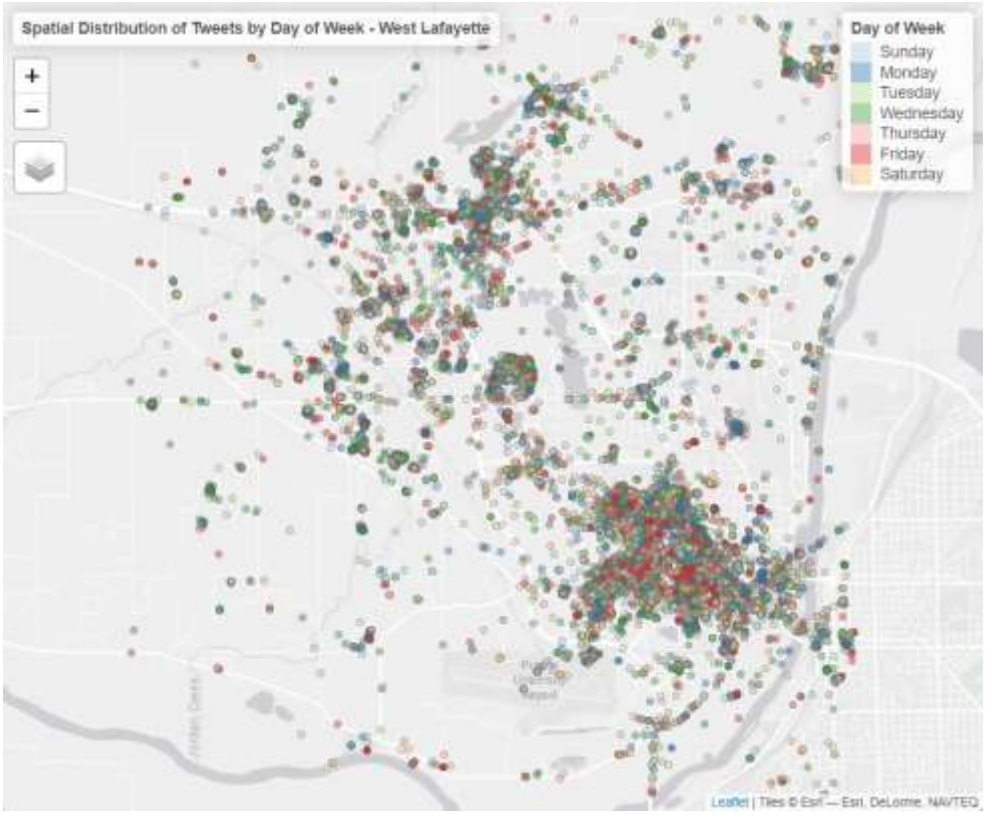


Lafayette area. The map showed that most of the tweet occurred around campus area (including on-campus resident halls), and the apartment area at the northwestern side of the main campus.

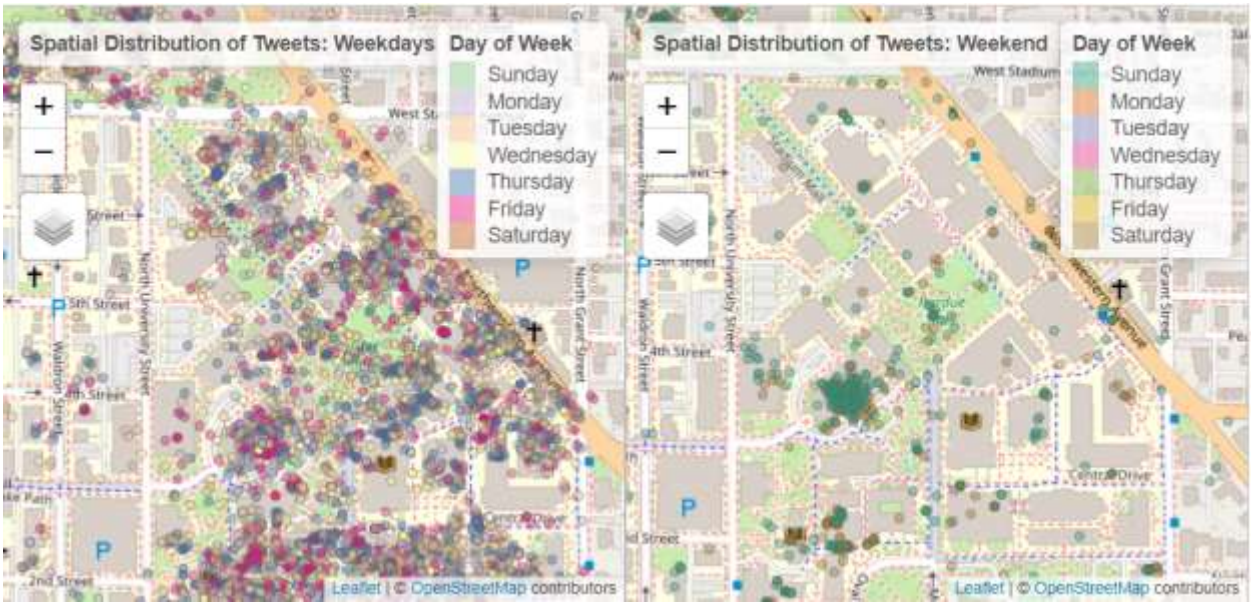
The next step for this part of the project is to provide monthly, weekly, and daily distribution of the tweets within West Lafayette area. The following map shows the distribution of tweets by month, notice that there are missing data during summer break (July, June, August) time.



The following maps represent the day of week distribution for the overall tweets. In this step, I have put weekdays, and weekend into two separate groups and tried to use “facet” function to plot two different layers in one map. The map showed that there are a significant amount of users tweeted on campus during the weekdays, and there are a lot less users tweeted on campus, during weekends.



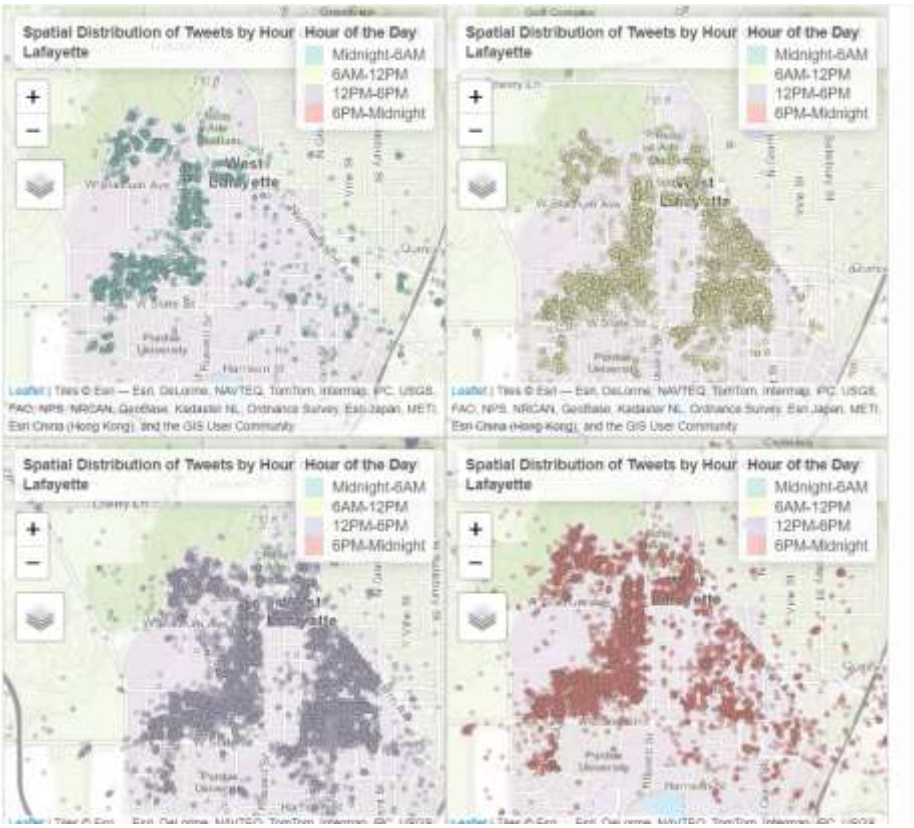
One Significant challenge I have faced is that the facet map legend showed all the days on each side of the layer. In other words, the layer of weekday also prints “Saturday” and “Sunday”, and vise



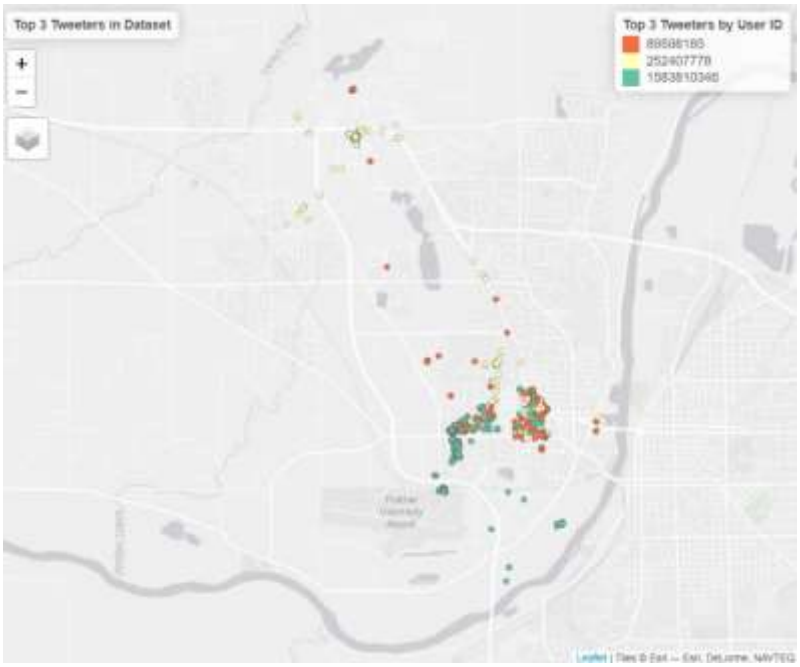
versa, however, if I delete the column function, the dots will not be colored (additional map in reference).

The third step of this part of the project is to conduct the hour distribution of the tweet data. I have split the data into four sections to have better visualization. In the code I have "Midnight to 6AM", "6AM to 12PM", "12PM to 6PM", and "12PM to Midnight". Each section has six hours. The first map I showed here represents the overall distribution of each time section of tweet. (in reference original map).

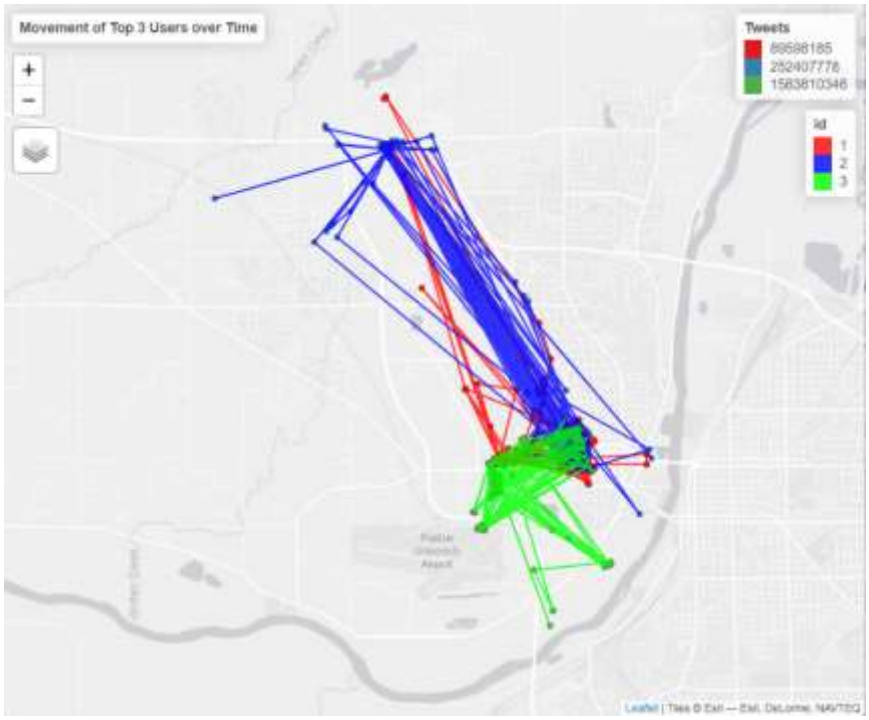
For the following map, I also tried to use the facet function and provide four sections of distribution into one map. The map showed that during the midnight, most of the tweet were at dorm and residence area. Department buildings (buildings with classrooms) have less tweets. In addition, during 6AM to 6PM, there are a lot of tweets generated near department buildings. At last, when we look at the 6PM to Midnight, there were still some tweets generated near department buildings.



The next step for this part of the project is to draw the distribution of the top three users. The map below shows the location of top three users' tweet location. From the map we can see that the first and the third most user generated most of the tweet on campus. In contrast, the yellow user (2^{nd}) had a lot of tweet generated outside the campus. One possible reason is that this user is living at the northwest side of campus. That place can be this users' apartment.

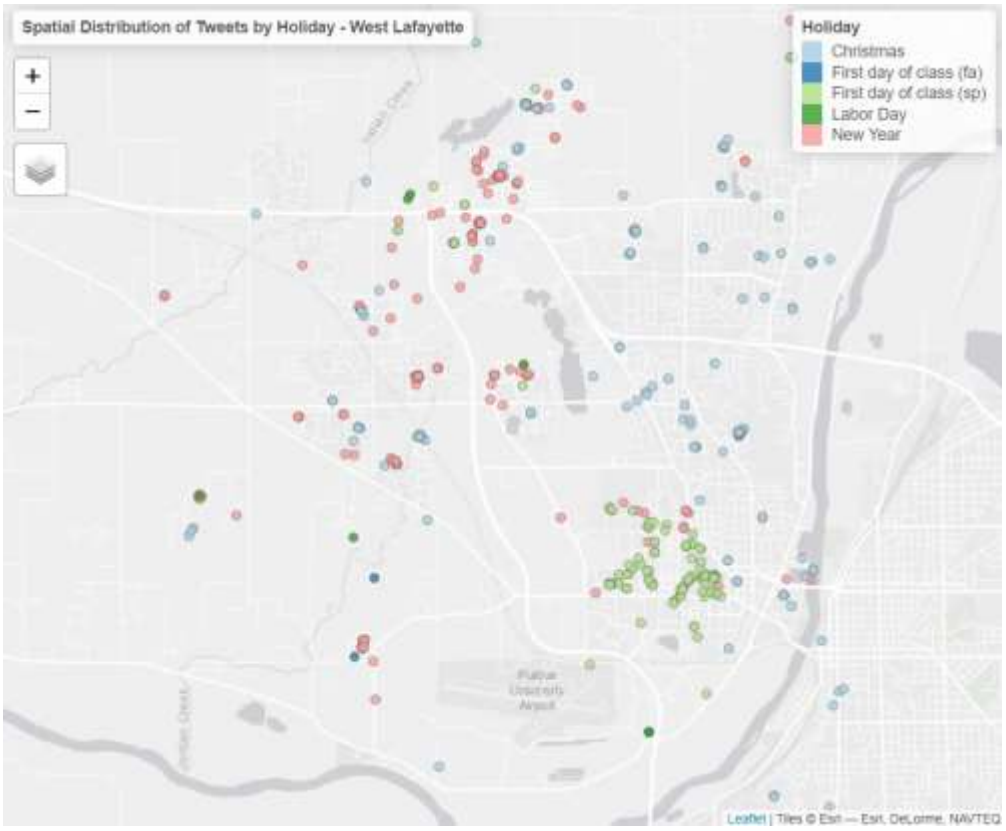


The next step is to provide the spatial temporal distribution of top three users. I have draw the moving patterns of top three users in the following map. The second users are have a lot of travels between his or her apartment and the campus.



5. Mapping the story of your own interest (AOI, individuals)

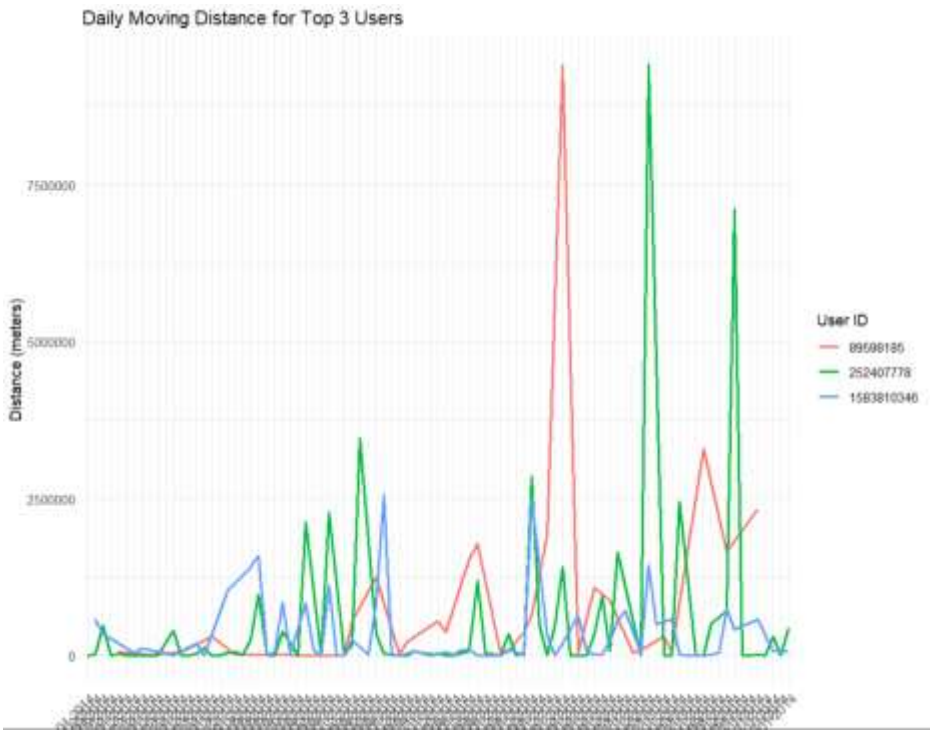
For this part of the project, I have selected some holidays and tried to use the map to show the distribution of tweets during different holidays. The map distribution showed the tweets of Christmas, Labor Day, New Year, First day of class (Fall 2014, and Spring 2014). There are more student tweet on campus during the first day of class as compared to holidays (Christmas and Ney Year). It shows that students were returned to campus for class to begin. Another interesting fact is that more people tweet at entertaining facilities, such as restaurant and some grocery stores during Christmas while most people tweet at home during New Year.



In addition, I have drawn a bar plot of showing the amount of tweet during these selected days, however, with the lack of data, there were less tweet during selected days in summer (chart in reference).

6. Mapping/estimate mobility

The next step is to conduct the daily moving distance of the top three users. The graph below shows the moving distance of each tweeter. The x axis of the map represents the day and the y axis represents the moving distance. While we can see that some users are not moving in certain days.

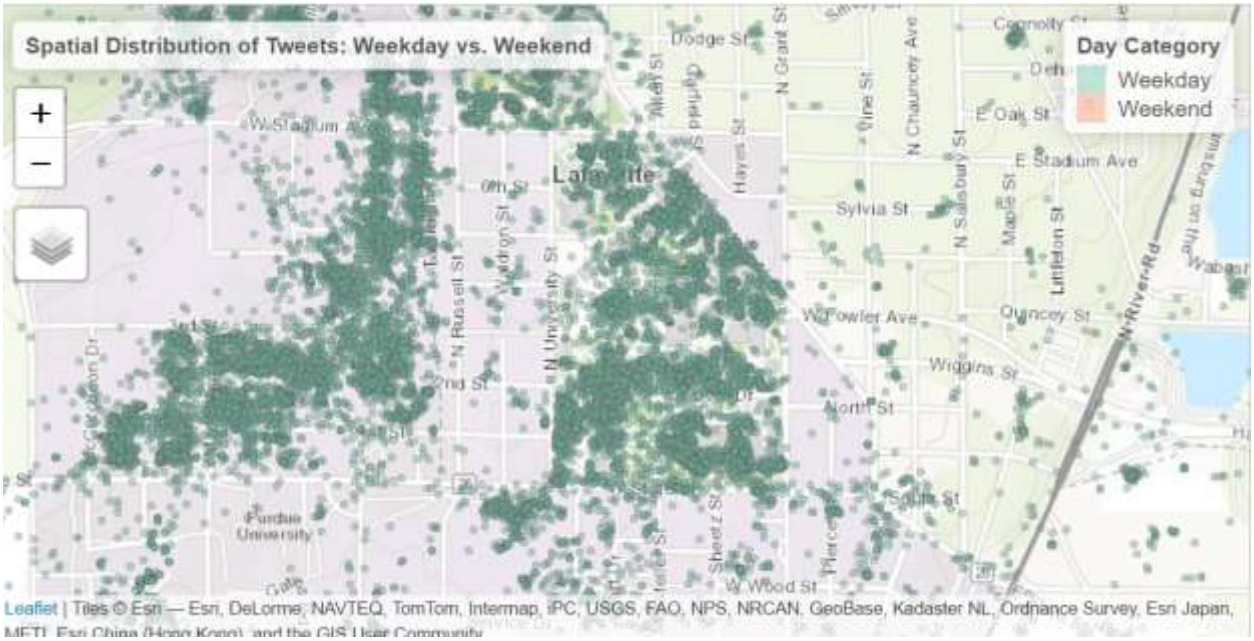


7. Summary/Conclusion/Concluding Remarks

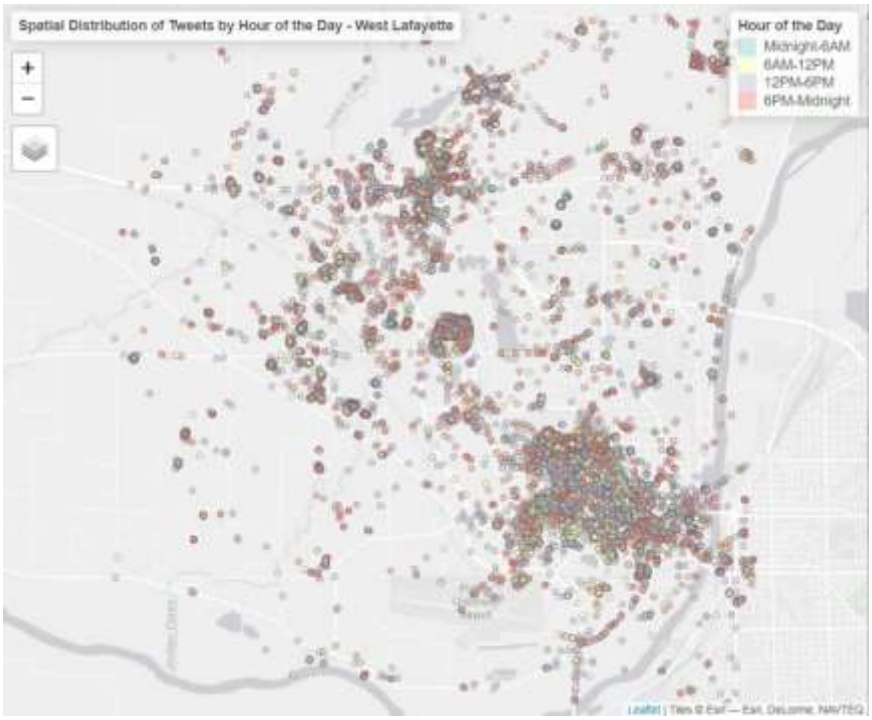
This challenge practiced the spatial temporal exploration of tweets. Through the analysis, we found that most of the tweets on weekdays occurred near department buildings and less tweets were generated near these areas. In addition, less tweets were generated near department buildings during midnight. Due to the missing data in summer, it is hard to inform the holiday difference while comparing it to other holidays (such as Christmas).

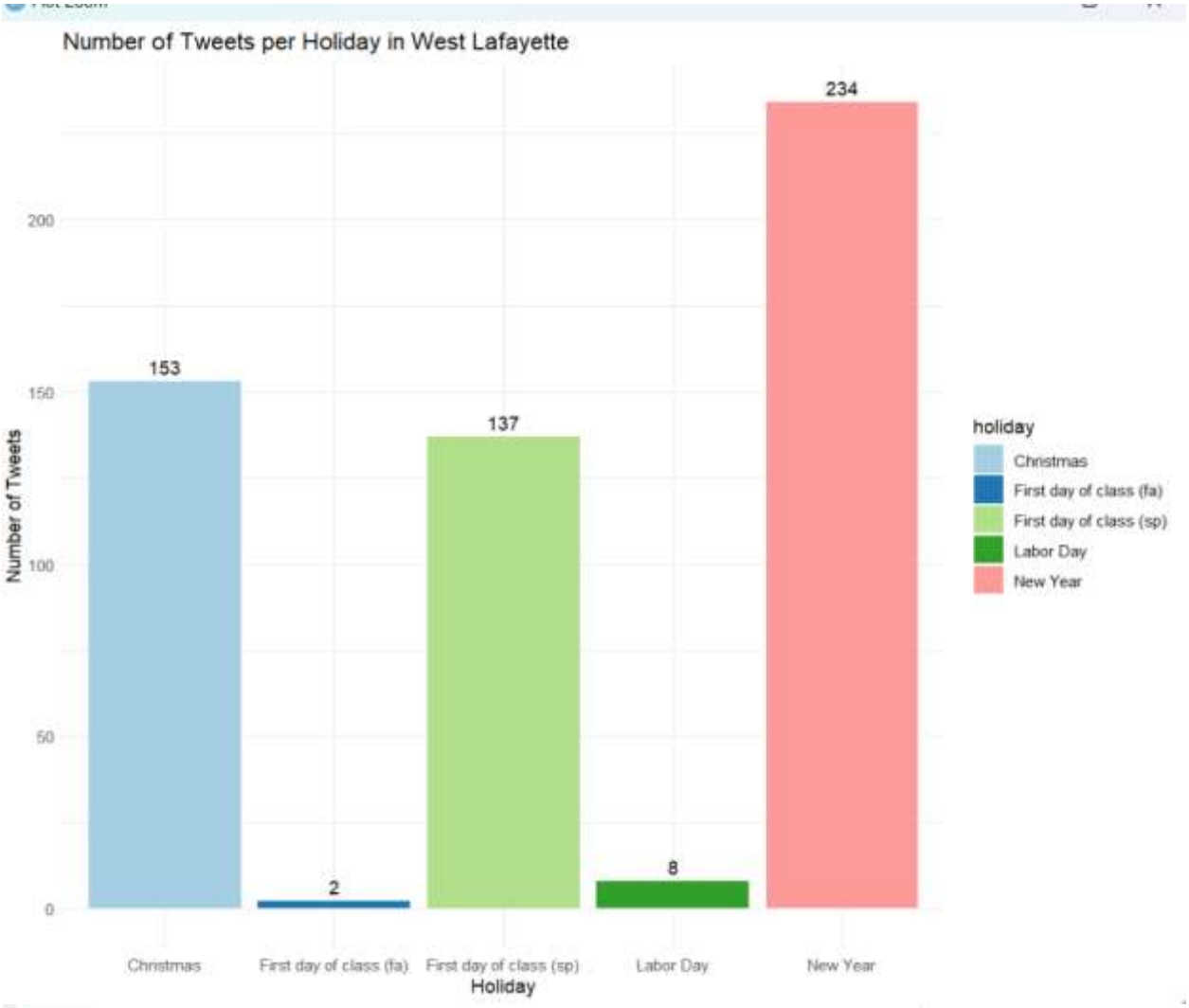
Acknowledgement

References


Spatial Distribution of Tweets: Weekday vs. Weekend
Day Category
Weekday
Weekend
Dodge St
E Oak St
E Stadium Ave
Sylvia St
W Fowler Ave
Quincey St
Wiggins St
North St
South St
W Wood St
Purdue University
N River Rd
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community


Spatial Distribution of Tweets: Weekday vs. Weekend
Day Category
Weekday
Weekend
West Lafayette
Dodge St
Connolly St
Dehart St
E Oak St
E Stadium Ave
W Stadium Ave
Sylvia St
W Fowler Ave
Quincey St
Wiggins St
North St
South St
W State St
Oval Dr
Purdue University
N River Rd
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community


Spatial Distribution of Tweets by Hour of the Day - West Lafayette
Hour of the Day
Midnight-6AM
6AM-12PM
12PM-6PM
6PM-Midnight




```
#extra
#create KML file
st_write(sf_read, "sf_kml.kml", driver = "KML", append = FALSE) #sf
writeOGR(obj = sp_read, dsn = "sp_kml.kml", driver = "KML", overwrite_layer = TRUE) #sp

#create object sf or sp from KML
sp_from_kml <- readOGR("sp_kml.kml")
sf_from_kml <- st_read("sf_kml.kml")

#same as above can try geopackage file
#create GeoPackage file
st_write(sf_read, "sf_geopackage.gpkg", driver = "GPKG", append = FALSE) #sf
writeOGR(obj = sp_read, dsn = "sp_geopackage.gpkg", driver = "GPKG", overwrite_layer = TRUE) #sp

#create object sf or sp from GeoPackage
sp_from_geopackage <- readOGR("sp_geopackage.gpkg")
sf_from_geopackage <- st_read("sf_geopackage.gpkg")
```