

HW6: Spatio-temporal data

Fern Bromley

December 5th, 2025

Los Angeles Bikeshare

For this lab, you will study an expanded version of the bikeshare data we saw in class. Instead of looking at a single day, we'll look at a whole week broken up into two-hour intervals.

I. Exploring the data [26 pts]

- (1) [6 pts] First load the `spacetime` and `sp` packages and then load the `departures_STFDF` object stored in `bike_share_HW6.RData`. What is the maximum number of bikes that departed a single station in any single two-hour window during the week of January 5-11, 2020? Where is the station at which this occurred? On what date and during which time interval did the maximum occur? (*Hint*: the spatial, temporal, and covariate information inside `departure_STFDF` can be accessed using `@sp`, `@time`, and `@data`, respectively.)

```
# install.packages("spacetime")
library(spacetime)
# install.packages("sp")
library(sp)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.5.2      v tibble     3.2.1
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.0.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
load('bike_share_HW6.RData')
```

```
max(departures_STFDF@data$count)
```

```
## [1] 37
```

```
departures_STFDF@data$station_id[which(departures_STFDF@data$count==max(departures_STFDF@data$count))]
```

```
## [1] 3030
```

```
station <- subset(departures_STFDF@data, departures_STFDF@data$station_id == 3030) %>%
  arrange(-count) # sort by count, descending
```

At station 3030 in the downtown area, 37 bikes were checked out at 4pm on Friday (Jan 10th). This was the most bikes checked out in any hour of the week at all stations.

- (2) [4 pts] What are the coordinates for the three stations that had the largest total number of departures over the one week period? How many total departures did each one have?

```
weekly_totals <- departures_STFDF@data %>%
  group_by(station_id) %>%
  summarise(departures = sum(count))
```

Stations 3030, 3005, and 3014 had the most total departures that week, with 361, 283, and 219 departures respectively.

- (3) [4 pts] Overlay all 187 bikeshare locations on a map of Los Angeles and mark the three locations you found in question 3. Include your map in your submission.

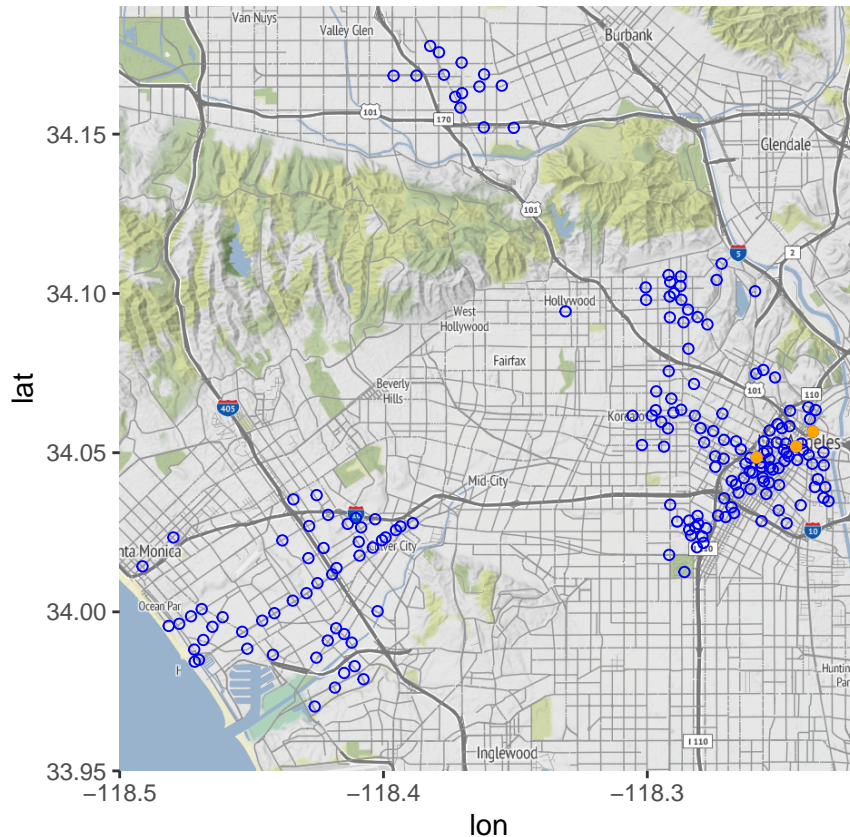
```
library(ggmap)
library(sf)

stations <- departures_STFDF@sp@coords
busy_stations <- subset(departures_STFDF,
  departures_STFDF@data$station_id %in% c(3030, 3005, 3014))@sp@coords

# st_bbox(departures_STFDF@sp)
la_map <- get_stadiamap(bbox = c(left = -118.5, right = -118.22,
  bottom = 33.95, top = 34.19),
  zoom = 12)
```

```
## i © Stadia Maps © Stamen Design © OpenMapTiles © OpenStreetMap contributors.
```

```
ggmap(la_map)+
  geom_point(data = stations, aes(x = lon, y = lat), color = "blue", shape = 1)+
  geom_point(data = busy_stations, aes(x = lon, y = lat), color = "orange")
```



- (4) [4 pts] What is the name of the major train station near these busy bikeshare stations? What is one other feature in the area that you think people might decide to bike to/from?

LA's Union Station is located near these bike stations. Government and municipal services like City Hall and court buildings are also nearby.

- (5) [4 pts] Add a new column to the data in `departures_STFDF` called `weekend` that has a value of `TRUE` for time intervals that fall on a Saturday or Sunday, and `FALSE` otherwise. Double check that you have $2 \times 12 \times 187 = 4488$ `TRUE` values and 11220 `FALSE` values.

```
departures_STFDF@data$weekend <- case_when(departures_STFDF@data$DoW == "Sat" ~ T,
                                             departures_STFDF@data$DoW == "Sun" ~ T,
                                             .default = F)
```

```
length(subset(departures_STFDF, departures_STFDF@data$DoW %in% c("Sat", "Sun"))) # i get 5040 trues, no
```

```
## [1] 5040
```

- (6) [4 pts] Do you expect the number of departures on the weekend to be higher or lower than the number during the week? Why?

I expect them to be higher during the weekend, since I assume most people commute by car and use the bikes for leisure/more casual trips.

II. A spatio-temporal model for counts [14 pts]

- (7) [4 pt] Create a new STFDF object called `departures_STFDF_proj` that projects the coordinates using zone 11 of the Universal Transverse Mercator coordinate system. Rename the coordinates `x` and `y` instead of `lon` and `lat`. SEE PDF ON GRADESCOPE. Provide executable code (**PEC**; hint: you'll need to use the projection function from the `sp` package, not `sf`).

```
#spTransform?

new_sp <- spTransform(departures_STFDF@sp, "epsg:32611")
coordnames(new_sp) <- c("x", "y")
departures_STFDF_proj <- STFDF(sp = new_sp, time = departures_STFDF@time, data = departures_STFDF@data)
```

- (8) [6 pts] Use the `mgcv` package to fit a GAM for a Poisson response with log link function in which weekend is a fixed effect. Model the spatio-temporal random effect using flexible basis functions of your choice over space and hour of the day. Make sure you use the projected spatial coordinates. **PEC**.

```
library(mgcv)
gam1 <- gam(count ~ s(x, y) + s(hour) + weekend, family = poisson, data = departures_STFDF_proj)
```

- (9) [4 pts] What is the point estimate for the effect of weekend? Does it agree with your expectations?

```
summary(gam1)

##
## Family: poisson
## Link function: log
##
## Formula:
## count ~ s(x, y) + s(hour) + weekend
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.48427    0.02382  -62.315   <2e-16 ***
## weekendTRUE   0.04969    0.02784   1.784    0.0744 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df Chi.sq p-value
## s(x,y)      28.676 28.985 3935.45 < 2e-16 ***
## s(hour)     6.172  7.334  20.31 0.00613 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) = 0.106   Deviance explained = 23.1%
## UBRE = 0.14097   Scale est. = 1          n = 15708
```

The estimate for the effect of weekends is positive, indicating that there were more generally bikes checked out during this weekend than compared to the week, which aligns with my prediction.