Progress Report Two

Noah Anderson

2023-10-21

Apendix

Setting Parameters and Loading Packages

```
library(RSocrata)
library(tidyverse)
library(sf)


# Define parameters for date range
start_date <- "2022-06-01"
end_date <- "2022-09-01"

# Define vector for community areas to be queried
target_ca <- c( "Uptown", "North Center", "Lincoln Square")
```

Importing Data

```
chicago_ca <- st_read("/Users/noahanderson/Documents/Data/Boundaries-Neighborhoods.geojson") %>%
  st_transform(4269) %>%
  filter(pri_neigh %in% target_ca)

divvy_stations_url <- "https://data.cityofchicago.org/resource/bk89-9dk7.json"
divvy_stations <- read.socrata(divvy_stations_url) %>%
  st_as_sf(coords = c("longitude", "latitude"), crs = 4269)

# Join data frames to identity stations in target community areas
divvy_stations_in_ca <- st_join(chicago_ca, divvy_stations)

# Define string for target station ID's
ca_ids_string <- paste(divvy_stations_in_ca$id, collapse = ",")

# Define the base URL for historical divvy time snapshots
divvy_time_base_url <- "https://data.cityofchicago.org/resource/eq45-8inv.json"

# Define the select, where, limit, and order parameters
select_columns <- "$select=id,timestamp,docks_in_service,available_docks"
id_filter <- paste0("id in (", ca_ids_string, ")")
```

```r
date_filter <- paste0("timestamp >= '", start_date, "' AND timestamp <= '", end_date, "'")
combined_filter <- paste0("$where=", id_filter, " AND ", date_filter)
order_parameter <- "$order=timestamp DESC"

# Combine the base URL with the select, where, limit, and order parameters
divvy_time_url <- paste0(divvy_time_base_url, "?", select_columns, "&", combined_filter, "&", order_para

# Read divvy data
divvy_time_raw <- read.socrata(divvy_time_url)
```

Cleaning Data

The final data set is put into a wide T * N data frame where the rows are time stamps (T) and the columns are nodes (N). In this case, each divvy bike station is an ID.

```r
# Clean divvy data
divvy_time_clean <- divvy_time_raw %>%
  mutate(deficit = as.numeric(available_docks) - as.numeric(docks_in_service)) %>%
  select(timestamp, id, deficit) %>%
  distinct(timestamp, id, .keep_all = TRUE)

# Pivot data to wide format
divvy_time_wide <- divvy_time_clean %>%
  pivot_wider(names_from = id, values_from = deficit) %>%
  select(-timestamp)

# Remove column names
colnames(divvy_time_wide) <- NULL

# Save csv to StemGNN sub repository
write.csv(divvy_time_wide, "/Users/noahanderson/Documents/GitHub/cap-stone-PDAT/src/StemGNN/dataset/div
          row.names = FALSE, col.names = FALSE)
```

References

RSocrata Devlin, H., Schenk Jr., T., Leynes, G., Lucius, N., Malc, J., Silverberg, M., & Schmeideskamp, P. (2023). "RSocrata: Download or Upload 'Socrata' Data Sets." R package version 1.7.15-1. Available at: https://CRAN.R-project.org/package=RSocrata

tidycensus Walker, K., Herman, M., & Eberwein, K. (2023). "tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames." R package version 1.5. Available at: https://CRAN.R-project.org/package=tidycensus

tidyverse Wickham H., Averick M., Bryan J., Chang W., McGowan L.D., François R., Grolemund G., Hayes A., Henry L., Hester J., Kuhn M., Pedersen T.L., Miller E., Bache S.M., Müller K., Ooms J., Robinson D., Seidel D.P., Spinu V., Takahashi K., Vaughan D., Wilke C., Woo K., Yutani H. (2019). "Welcome to the tidyverse." Journal of Open Source Software, 4(43), 1686. doi:10.21105/joss.01686.

sf Pebesma, E. (2018). "Simple Features for R: Standardized Support for Spatial Vector Data." The R Journal, 10(1), 439–446. doi:10.32614/RJ-2018-009. Available at: https://doi.org/10.32614/RJ-2018-009

riem Salmon, M. [aut, cre], Anderson, B. [rev], CHAI Project [fnd], rOpenSci [fnd], & Herzmann, D. [ctb] (2022). "riem: Accesses Weather Data from the Iowa Environment Mesonet." R package version 0.3.0. Available at: https://CRAN.R-project.org/package=riem

lwgeom Pebesma, E., Rundel, C., Teucher, A., & liblwgeom developers (2023). "lwgeom: Bindings to Selected 'liblwgeom' Functions for Simple Features." R package version 0.2-13. Available at: https://CRAN.R-project.org/package=lwgeom

Steif, K. (2021). Public Policy Analytics: Code and Context for Data Science in Government. CRC Press.