Weather and Rideshare Ridership

John Cruz

2023-04-25

Data Source

Weather (Oikolab)

Data was collected using Oikolab API historical data API service. It collects its data from the ECWMF and NOAA.

Uber & Lyft Trips (NYC Taxi and Limousine Commission)

Data was collected using the available Parquet File. The agency collects the data from Uber and Lyft.

Data Preparation

Required Libraries

```
library(tidyverse)
library(arrow)
library(weathermetrics)
library(lubridate)
```

Load Historical Weather Data

```
weather <- read_csv('oikolabs.csv') |>
  janitor::clean_names()
```

Title

Details

```
weather <-
  weather |>
  mutate(temp_deg_f = celsius.to.fahrenheit(temperature_deg_c),
         rel_humidity = dewpoint.to.humidity(t = temperature_deg_c,
                                             dp = dewpoint_temperature_deg_c,
                                             temperature.metric = "celsius"),
         heat_idx = heat.index(t = temp_deg_f,
                               rh = rel_humidity))
```

```r
weather_trim <-
  weather|>
  select(datetime_utc, temp_deg_f, rel_humidity, heat_idx)

weather_trim <-
  weather_trim |>
  mutate(day_of_week = wday(datetime_utc, label = TRUE, week_start = 1, abbr = FALSE))
```

```r
tlc_trips <- read_parquet('fhvhv_tripdata_2022-08.parquet')

tlc_trips_trim <-
  tlc_trips |>
  select(hvfhs_license_num, pickup_datetime, dropoff_datetime, PULocationID, DOLocationID, trip_miles,
```

Research question

Does high heat index days (>=90) increase the number of trips taken with Uber or Lyft compared to non-high heat index days?

Cases

Weather

Each case represents hourly weather measurements in August 2022. There are a total of 2,184 observations.

Uber & Lyft Trips

Each case represents a trip taken either via Uber or Lyft. There are a total of 17,185,687 observations in the month of August 2022.

Type of study

This is an observational study.

Dependent Variable

The response variable is total trips and is numerical

Independent Variable(s)

The independent variables are:

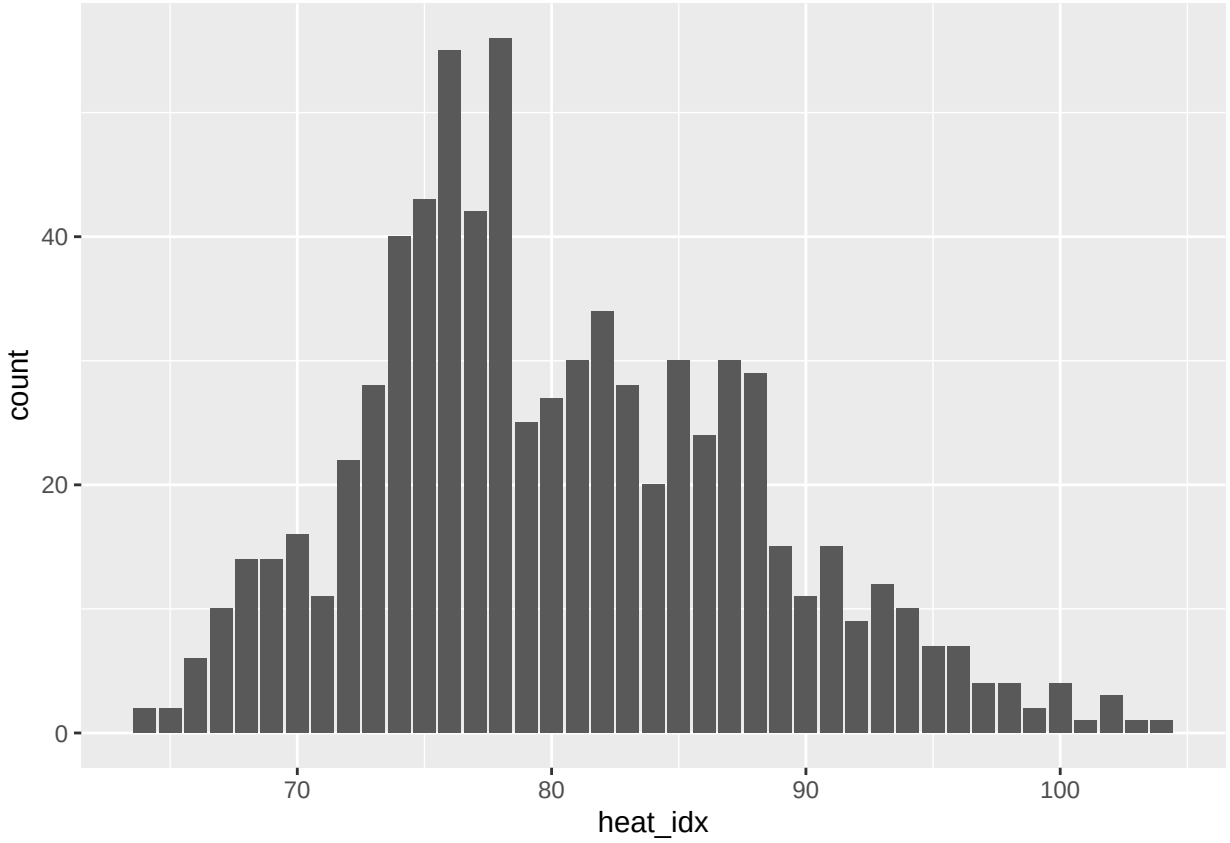
- heat index: numerical
- day_of_week: categorical

Note: Other potential factors that are important but not included: precipitation, special events (i.e. sporting event), major delays with public transportation (MTA Subway)

Relevant summary statistics

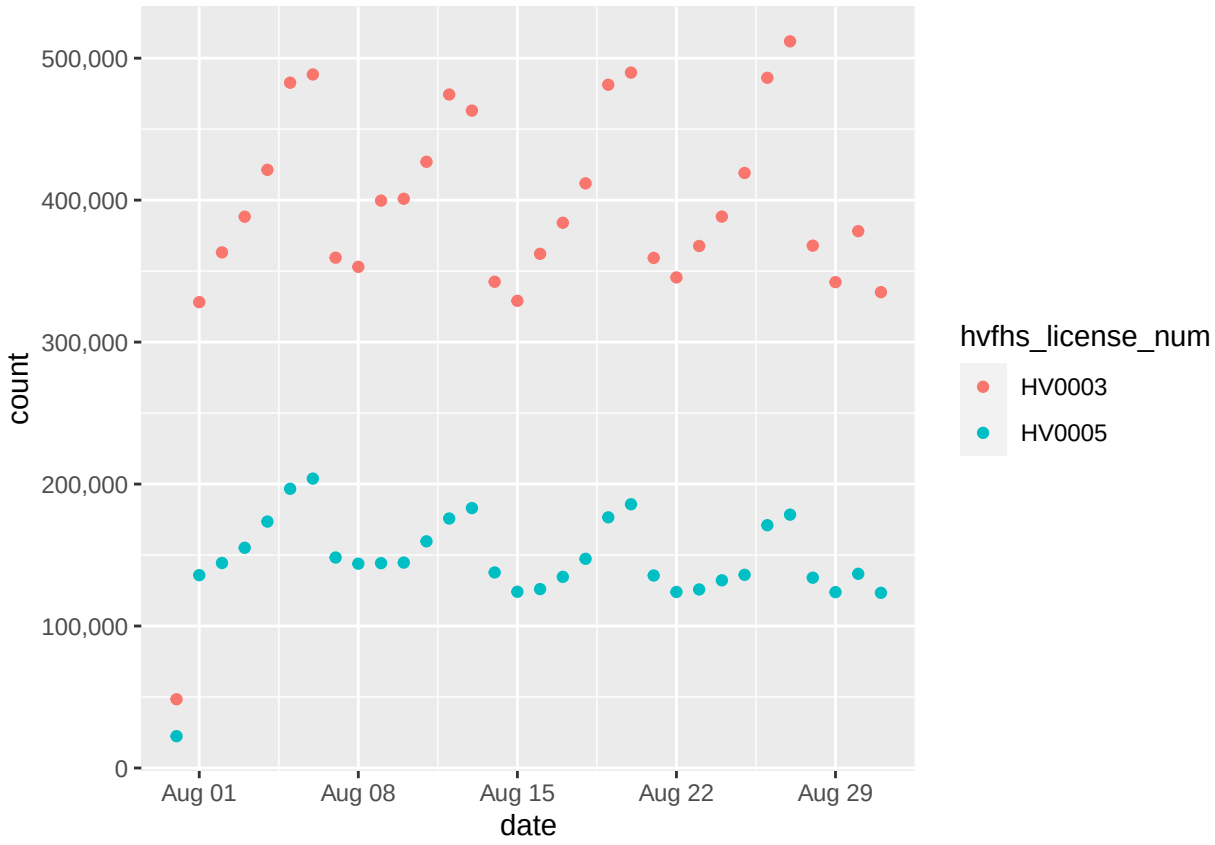
Count of hours for heat index throughout the month of August 2022.

```
monthly <-
  weather_trim |>
  mutate(month = month(datetime_utc), date = date(datetime_utc)) |>
  filter(month == 8) |>
  group_by(heat_idx) |>
  summarise(count = n())

monthly |>
  ggplot(aes(x = heat_idx, y = count)) +
  geom_bar(stat = 'identity')
```



Count of trips by Uber and Lyft in August.

```
trip_counts <-
  tlc_trips_trim |>
  mutate(date = date(pickup_datetime)) |>
  group_by(hvfhs_license_num, date) |>
  summarise(count = n(), .groups='keep')

trip_counts |>
  ggplot(aes(x = date, y = count, colour = hvfhs_license_num)) +
  geom_point(stat = 'identity') +
  scale_y_continuous(labels = scales::comma)
```




```
tlc_trips_trim |>
  mutate()
```

```
## # A tibble: 17,185,687 x 8
##    hvfhs_licen~1 pickup_datetime     dropoff_datetime    PULoc~2 DOLoc~3 trip_~4
##    <chr>         <dttm>              <dttm>                <int>   <int>   <dbl>
##  1 HV0003        2022-07-31 17:03:30 2022-07-31 17:14:08      35      61    2.59
##  2 HV0003        2022-07-31 17:31:21 2022-07-31 18:08:01      65      39    7.59
##  3 HV0003        2022-07-31 17:26:14 2022-07-31 17:39:20     188      89    2.49
##  4 HV0003        2022-07-31 17:02:24 2022-07-31 17:23:37     237     243    9.21
##  5 HV0003        2022-07-31 17:41:52 2022-07-31 17:55:53     239      90    3.11
##  6 HV0003        2022-07-31 17:01:26 2022-07-31 17:15:44     114     158    1.06
##  7 HV0003        2022-07-31 17:24:31 2022-07-31 17:31:05     158      68    1.68
##  8 HV0003        2022-07-31 17:06:27 2022-07-31 17:28:36     138      10    9.88
##  9 HV0003        2022-07-31 17:54:47 2022-07-31 18:06:49     244     119    1.81
## 10 HV0003        2022-07-31 17:17:44 2022-07-31 17:51:41     132     265   17.1
## # ... with 17,185,677 more rows, 2 more variables: trip_time <int>,
## #   base_passenger_fare <dbl>, and abbreviated variable names
## #   1: hvfhs_license_num, 2: PULocationID, 3: DOLocationID, 4: trip_miles
```