# Libraries Used

The packages used in this analysis are the same from the prior analysis, Tidyverse for data manipulation, sf for modifying spatial data, tigris for getting the basemaps to plot my routes and extrafont to bring in new fonts for the plots.

library(tidyverse) # Data Manipulation library(sf) # Manipulation Spatial Data

library(tigris) # Getting Tract and Roads Spatial Data library(extrafont) # Better Fonts For GGPLOT

# Data Used

The data is also the same running route data from the prior post.

runs\_and\_routes <- readRDS('data/runs\_and\_routes.RDS') all\_routes <- readRDS('data/all\_routes.RDS')

For the basemap I’m again using the tigris package however this time getting census tracts rather than roads. According to the package, *Census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people*. The map is downloaded using the tracts() function with inputs for state and county.

nyc\_tracts <- tracts("NY", "New York", cb = T) %>% st\_transform(crs = st\_crs(runs\_and\_routes$geometry))

ggplot() + geom\_sf(data = nyc\_tracts) + ggthemes::theme\_map()



Unlike the analysis where the heatmap was just overlaid atop the map, here I need to identify which census tracts contained a route I ran vs. which didn’t. This can be done using the st\_join function, specifying it to be a left join, and specifying the join type as st\_intersects which joins the route information if the lat/long is contained in the census tract. The data is then grouped by tract\_name and some other tract metadata. Then I create a field for the number of routes contained in each census tract, which will be used for the choropleth.

#Join Routes to Tracts by Intersecting nyc\_geo\_join <- nyc\_tracts %>%

st\_join(all\_routes %>% distinct(route\_id, geometry), join = st\_intersects,

left = T

) %>%

group\_by(

TRACTCE, #Census Tract ID ALAND, #Land Area

AWATER #Water Area

) %>%

summarize(num\_routes = n\_distinct(route\_id, na.rm = T), .groups = 'drop') %>%

#Set 0 Routes to NA colored

mutate(num\_routes = if\_else(num\_routes == 0, NA\_integer\_, num\_routes))

# Visualization

The choropleth provides an alternative version to the heatmap which will better show each census tract that **at least one** of my routes had passed through. Really rare routes did not show up on the heatmap, but they will be clearer here.

ggplot() +

geom\_sf(data = nyc\_geo\_join, aes(fill = num\_routes)) +

scale\_fill\_viridis\_c(na.value = "grey90", guide = F) + coord\_sf(xlim = c(-74.15, -73.8)) +

labs(title = paste0("Census Tracts I've ",emo::ji('running')," Through"),

fill = "# of Routes Run", caption = "\*\*Author:\*\* JLaw") +

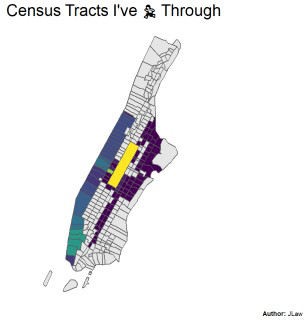
ggthemes::theme\_map() + theme(

plot.title = element\_text(size = 18, family = 'Arial Narrow', hjust

= .5),

plot.caption = ggtext::element\_markdown(), plot.caption.position = 'plot'

)



Now the East Side routes are clearer.

# What % of Manhattan Did I Run Through?

The island of Manhattan covers 22.7 square miles. I was curious what % of square miles I covered based on census tracts. While this will seriously over-count my distance covered it is easy to calculate. If I ran through the tract I get to count 100% of its land area. If I did not, I count nothing.

The ALAND columns from the Census Tract data contains the land area in square kilometers which I convert to square miles.

data\_summary <- nyc\_geo\_join %>% as\_tibble %>%

mutate(covered = !is.na(num\_routes)) %>% group\_by(covered) %>%

summarize(tracts = n(),

#Convert Square KM to Square Miles area = sum(ALAND)/2589988) %>%

mutate(pct\_tracts = tracts / sum(tracts), pct\_area = area/sum(area))

During this marathon training, I ran through 101 of Manhattan’s 288 Census Tracts (35%) and passed through census tract’s covering 8.7 *mi^2* out of 22.7 *mi^2* for **38.4%**.