Final Paper: Loopholes in Local Tobacco Policies

School of Global Policy and Strategy
GPEC 443 GIS and Spatial Data Analysis
Fall 2024

Abstract

This project aimed to investigate the relationship between local tobacco flavor restrictions and the prevalence of e-commerce tobacco retailers, while additionally focusing on Social Deprivation Index scores across San Diego County municipalities. Key research questions include whether brick and mortar tobacco shops are more likely to use e-commerce in areas with flavor bans, and what SDI characteristics are associated with higher concentrations of e-commerce retailers. Using pilot data from the Tobacco E-commerce Lab[1] at UCSD, findings aim to inform further research and improve regulatory frameworks for flavored tobacco products.

[1] The Tobacco E-commerce Lab: https://www.tobaccoecommercelab.com/

Background

E-cigarettes contain the highly addictive neurotoxin nicotine and emit aerosols with at least ten Proposition 65-listed chemicals known to cause cancer, birth defects, or reproductive harm.[2] They increase oxidative stress, a key factor in toxicity and addiction, and contribute to health issues like asthma, depression, and social adjustment challenges in young adults.[3] Flavored tobacco products have driven youth tobacco use, with 85.6% of youth tobacco users reporting flavored product use.[4] Thus, enforcing restrictions on addictive products and underage sales is critical.

California's Tobacco Endgame strategy aims to eliminate tobacco industry harm by 2035, partly through enforcing flavor restrictions. Governor Newsom's SB-793[5], banning most flavored tobacco products, faced a referendum-backed delay until 2022. Meanwhile, 108 local jurisdictions enacted flavor bans, ensuring some protections. However, SB-793 excluded e-commerce platforms, creating a loophole for online sales. This shift toward e-commerce by previously brick-and-mortar establishments undermines restrictions. Following SB-793, online searches for cigarettes and vapes spiked by 194% and 162%, respectively and preliminary data suggest that tobacco retailers in jurisdictions with flavored tobacco product bans are 3.3 times more likely to operate online[6], highlighting potential gaps in youth protection .

Youth surveys indicate increased online purchasing of flavored products despite bans.[7] Surveillance of e-commerce is vital to enforce compliance. California's AB-3218[8], effective January 1, 2025, will further regulate online sales, banning flavored product deliveries to limit youth access. This is significant, as youth accounted for 40% of online vape purchases in 2021.[9] Marginalized youth face greater nicotine addiction risks due to targeted tobacco marketing.[10] Incorporating variables like poverty, education, and housing insecurity, the Social Deprivation Index[11] quantifies geographic disadvantage, with higher scores indicating greater deprivation.

My questions include whether flavor bans increase e-commerce use by tobacco retailers, and how SDI scores correlate with e-commerce prevalence. I hypothesized that local bans drive retailers to e-commerce and that more deprived regions face higher tobacco e-commerce prevalence.

---

[2] California Department of Public Health, 2015. *California youth tobacco survey 2023 annual report*.
[3] Gunnell, Appleby, & Araya, 2019. The contribution of social determinants of health to suicide rates: An overview of the evidence.
[4] California Department of Public Health. (2023). *California youth tobacco survey 2023 annual report.*
[5] Bill Text - SB-793: Flavored tobacco products.
[6] Leas et al., 2023. E-commerce licensing loopholes: A case study of online shopping for tobacco products following a statewide sales restriction on flavoured tobacco in California.
[7] Chaffee et al., 2024. Flavored tobacco product use among California adolescents before and immediately after a statewide flavor ban.
[8] Bill Text - AB-3218: Unflavored Tobacco List.
[9] Do et al., 2023. Underage Youth Continue to Obtain E-Cigarettes from Retail Sources in 2022: Evidence from the Truth Continuous Tracking Survey.
[10] Tercyak et al., 2020. Prevalence and correlates of lifetime e-cigarette use among adolescents attending public schools in a low-income community in the US.
[11] The Robert Graham Center, n.d. *Social deprivation index*.

Methods

Data Sources & Measures: Retailer data, collected by The Tobacco E-commerce Lab at UCSD, includes a sample of all current tobacco retailers within the county of San Diego, identified through map-based searches using the Google Places API and Yelp. This data was then filtered with a focus on retailers offering e-commerce platforms. The resultant retailers are referred to as “brick-and-click” retailers, due to them having both online and in person means of sales. This particular data was collected in a pilot study conducted by the lab, who are now replicating this within the entire state of California.

The Social Deprivation Index (SDI) was used to examine socioeconomic factors potentially influencing tobacco e-commerce sales. US census tracts provided geometries for SDI scores to be geospatially merged to. This was then again merged, this time with the municipality names, retailer counts, and flavored tobacco restriction (FTR) status. Policy data indicating whether municipalities have FTRs were provided by the American Nonsmokers’ Rights Foundation, and I checked online with outside sources to ensure up to date reportings. Two main tables were used in my analyses, one aggregated on the retailer level, and one aggregated on the municipality level.

Data at Brick-and-click Retailer Level:

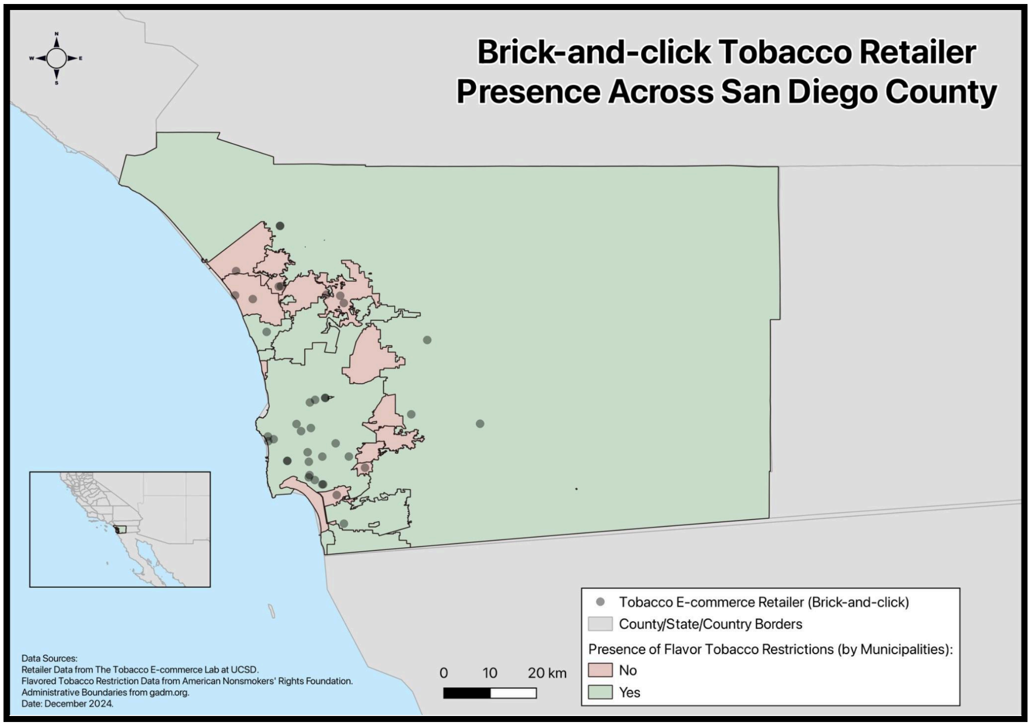
	lat	lon	name	FTR	geometry
1	32.69622	-117.1278	SAN DIEGO	1	POINT (-117.1278 32.69622)
2	32.71230	-117.1608	SAN DIEGO	1	POINT (-117.1608 32.7123)
3	32.69607	-117.1298	SAN DIEGO	1	POINT (-117.1298 32.69607)
4	32.60582	-117.0802	CHULA VISTA	1	POINT (-117.0802 32.60582)
5	33.02791	-116.8893	S.D. COUNTY	1	POINT (-116.8893 33.02791)
6	32.82564	-117.1559	SAN DIEGO	1	POINT (-117.1559 32.82564)
7	33.11287	-117.0805	ESCONDIDO	0	POINT (-117.0805 33.11287)
8	32.89485	-117.1230	SAN DIEGO	1	POINT (-117.123 32.89485)
9	32.89485	-117.1230	SAN DIEGO	1	POINT (-117.123 32.89485)

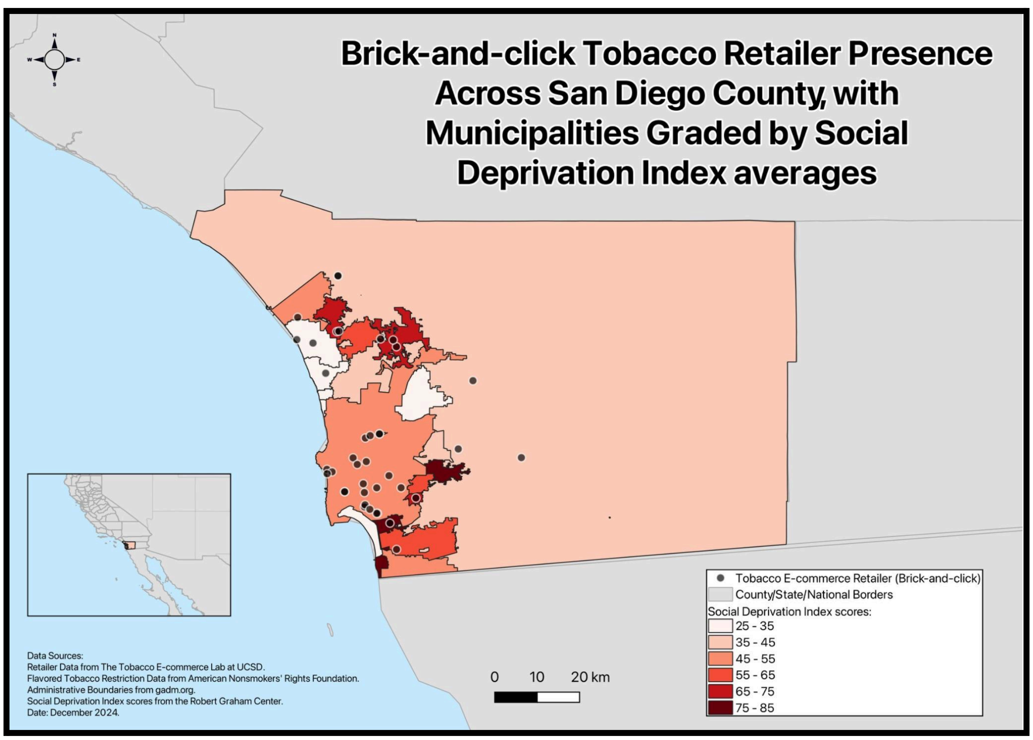
...

Data Aggregated on Municipality Level:

	name	geometry	FTR	retailers	average_sdi_score	Population
1	CARLSBAD	*POLYGON ((-117.2601 33.1571...*	0	2	26.58824	113018
2	CHULA VISTA	*POLYGON ((-116.9333 32.6483...*	1	1	62.43902	273841
3	CORONADO	*POLYGON ((-117.1186 32.5912...*	0	*NA*	30.75000	*NA*
4	DEL MAR	*POLYGON ((-117.2562 32.9795...*	0	*NA*	25.00000	*NA*
5	EL CAJON	*POLYGON ((-116.9622 32.8251...*	0	*NA*	76.33333	101963
6	ENCINITAS	*POLYGON ((-117.2166 33.0760...*	1	1	33.23077	60455
7	ESCONDIDO	*MULTIPOLYGON (((-117.0648 3...*	0	2	69.58333	147158
8	IMPERIAL BEACH	*POLYGON ((-117.1189 32.5349...*	1	*NA*	83.33333	25243
9	LA MESA	*POLYGON ((-116.9881 32.779,...*	0	*NA*	63.83333	60350
10	LEMON GROVE	*POLYGON ((-117.0231 32.7184...*	0	1	68.83333	26420
11	NATIONAL CITY	*POLYGON ((-117.0803 32.6928...*	0	1	84.83333	54910
12	OCEANSIDE	*MULTIPOLYGON (((-117.1688 3...*	0	1	53.12500	168705
13	POWAY	*POLYGON ((-117.0667 32.9749...*	0	*NA*	28.58333	47805
14	S.D. COUNTY	*MULTIPOLYGON (((-117.1063 3...*	1	5	39.66067	542063
15	SAN DIEGO	*MULTIPOLYGON (((-116.9308 3...*	1	20	52.52330	1388996
16	SAN MARCOS	*MULTIPOLYGON (((-117.1397 3...*	0	1	57.90909	93967
17	SANTEE	*POLYGON ((-116.9577 32.8580...*	0	*NA*	37.72727	59349
18	SOLANA BEACH	*POLYGON ((-117.2428 33.0074...*	1	*NA*	27.00000	*NA*
19	VISTA	*POLYGON ((-117.2724 33.2057...*	0	2	65.12500	98289

“*NA*” indicates that the given municipality does not have any tobacco e-commerce retailers present.





These maps created in QGIS mark the geolocations of brick-and-click retailers across San Diego County. In the upper map, municipalities are represented according to their flavored tobacco ban status. The lower map represents municipalities by their corresponding census tracts' average Social Deprivation Index scores.

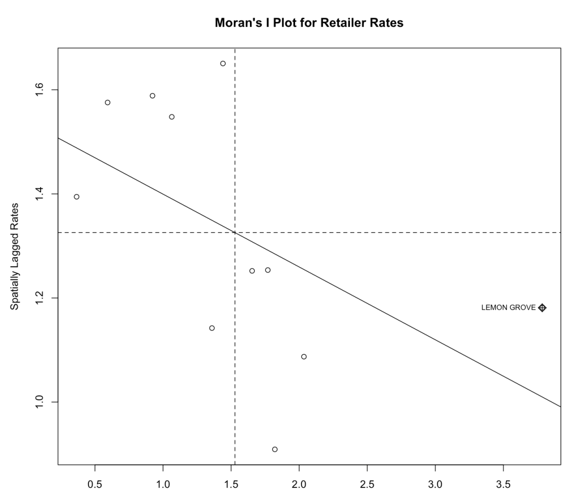
Analyses

To start my analyses, reprojection was necessary to align each of my datasets to the same coordinate reference system (CRS). I then conducted the Welch Two Sample t-test to compare the mean number of retailers per 100,000 population between municipalities with flavored tobacco restrictions (FTR = 1) and those without (FTR = 0). The test yielded a t-value of 1.4172 with 8.9714 degrees of freedom, and a p-value of 0.1902. The 95% confidence interval for the difference in means ranged from -0.4058477 to 1.7655263. These results indicate that the difference in the mean number of retailers per 100,000 population between the two groups is not statistically significant ($p > 0.05$). Specifically, the mean for municipalities without flavored tobacco restrictions (FTR = 0) was 1.775237, while the mean for municipalities with restrictions (FTR = 1) was 1.095397. This threw me off from my initial expectations, but I proceeded with additional analyses.

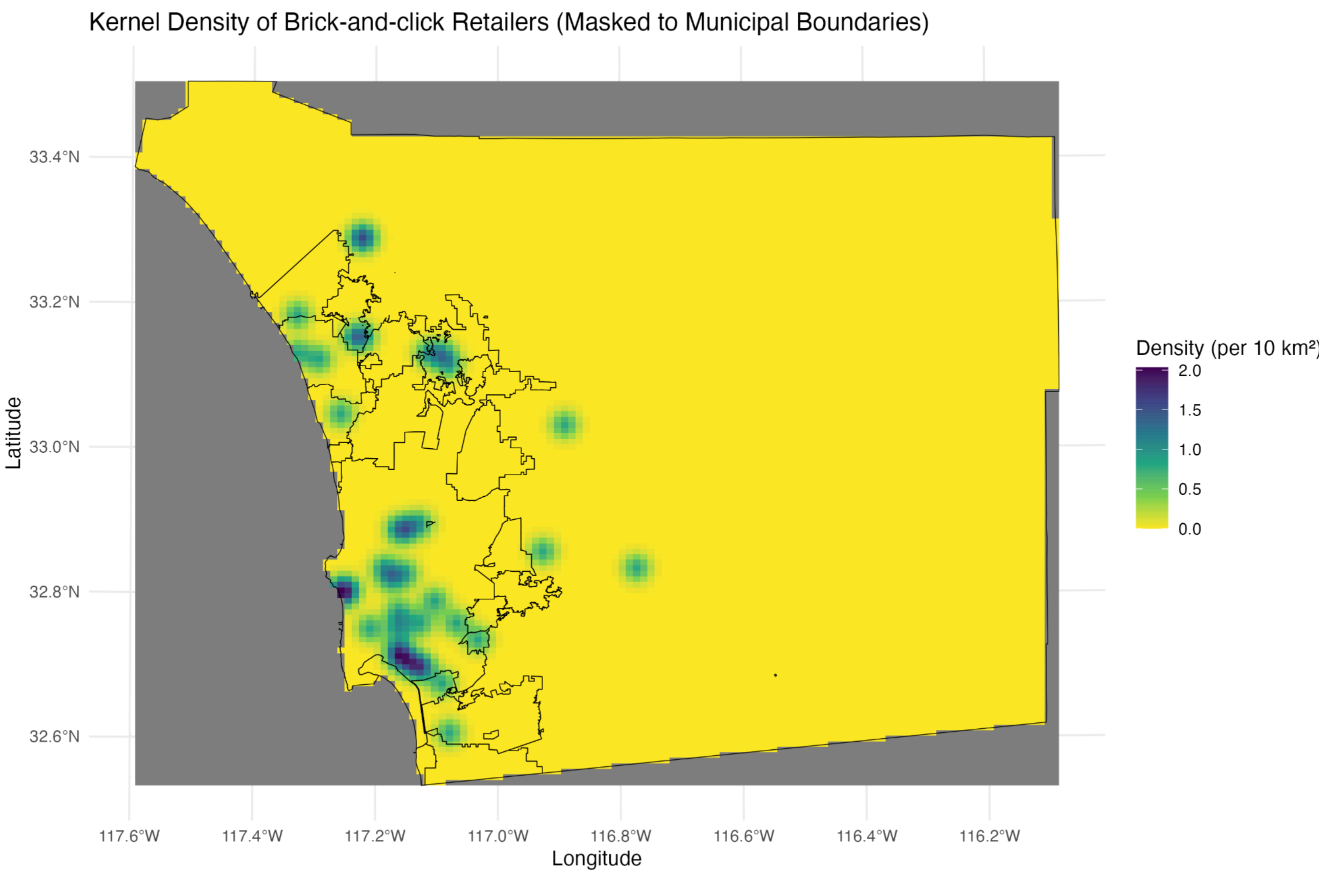
Comparing Brick-and-click Retailer Presence (Count Per 100,000 Population) in Municipalities with and without Flavored Tobacco Restrictions.

```
        Welch Two Sample t-test

data:  retailers_per_100k by FTR
t = 1.4172, df = 8.9714, p-value = 0.1902
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -0.4058477  1.7655263
sample estimates:
mean in group 0 mean in group 1
       1.775237        1.095397
```

A nearest neighbor analysis was conducted to examine the spatial distribution of brick-and-click tobacco retailers across municipalities. The observed mean nearest neighbor distance was 3.29 kilometers. This yielded a Nearest Neighbor Index of 0.6521, indicating a clustered spatial pattern, like we've observed in the visualizations. This indicated that there is some combination of factors influencing where e-commerce is being turned to, perhaps for regulatory circumvention. I also created a Moran's scatterplot which examines the spatial autocorrelation of retailer rates. The negative slope indicates a negative spatial relationship, where municipalities with high retailer rates tend to have neighbors with lower rates, and vice versa. Lemon Grove, identified as a high-low outlier, exhibits significantly higher retailer rates compared to its neighbors, warranting further investigation into its unique spatial pattern and into the particular tobacco e-commerce trends of Lemon Grove.



A kernel density estimation (KDE) was performed to visualize the density of retailers across the region. This KDE used a bandwidth of 1420 meters, smoothing the retailer points to highlight areas of concentration. The density values, originally calculated per square meter, were converted to represent densities per 10 square kilometers to better align with regional-level analysis. The KDE raster was then masked to municipal boundaries, ensuring that density values were restricted to relevant municipal areas. The resulting map presents density levels, with the color scale adjusted to represent densities between 0 and 2 retailers per 10 square kilometers. High-density areas are more concentrated around urban centers such as San Diego City, as indicated by darker purple marks. Brick-and-click retailers are less densely situated in eastern and northern regions of San Diego County.



The next step was to use R to to craft a map displaying OLS regression residuals, in an analysis that explored the effects of (1) flavored tobacco restrictions, (2) Social Deprivation Index scores, and (3) population size, on brick-and-click retailer counts across the municipalities of San Diego County.

Results of the regression model showed that population size was a highly significant predictor ($p < 0.001$), with a positive relationship to retailer counts ($\beta = 1.51\times10^{-5}$, $t = 12.15$). The model suggests that an increase of 100,000 people in the population corresponds to an expected increase of $1.51\times10^{-5} \times 100{,}000 = 1.512$ additional brick-and-click retailers. Conversely, flavored tobacco restrictions ($p = 0.127$) and SDI scores ($p = 0.764$) were not

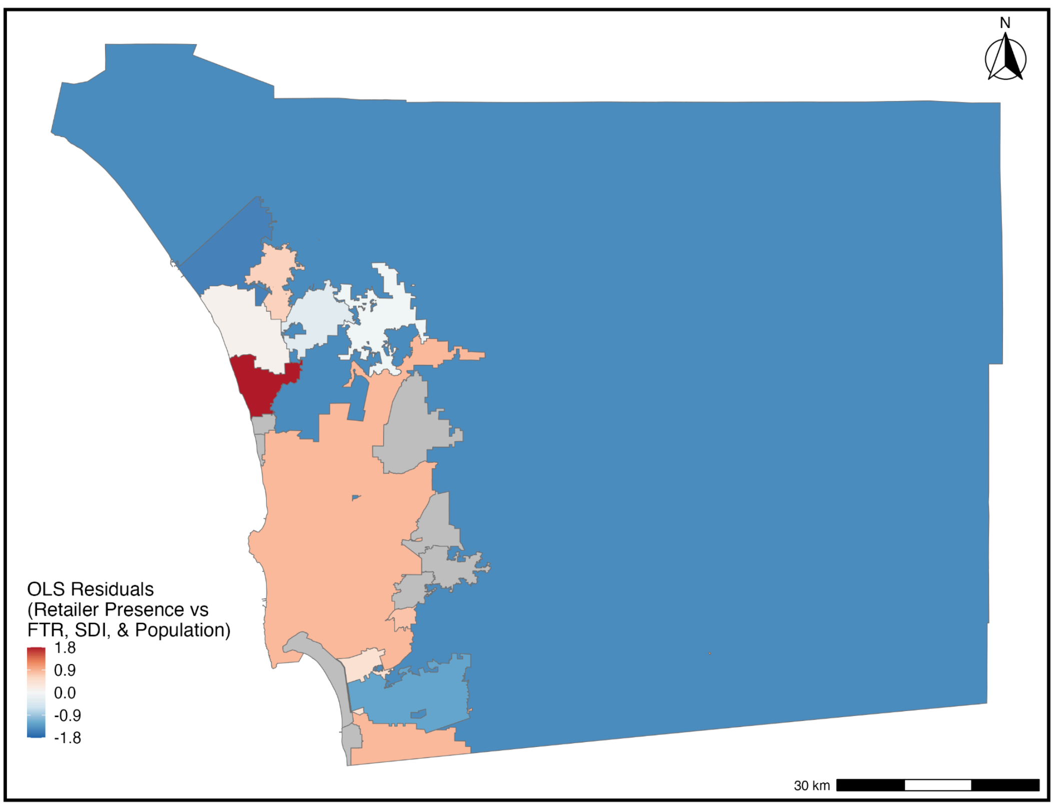
```
Call:
lm(formula = retailers ~ FTR + sdi_indice + Population, data = analyze)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5235 -0.7795  0.1087  0.6946  1.7925

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.845e-02  5.367e-01  -0.034    0.974
FTR         -1.836e+00  1.060e+00  -1.732    0.127
sdi_indice  -1.606e-01  5.147e-01  -0.312    0.764
Population   1.512e-05  1.245e-06  12.152 5.84e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.259 on 7 degrees of freedom
Multiple R-squared:  0.9652,	Adjusted R-squared:  0.9502
F-statistic: 64.63 on 3 and 7 DF,  p-value: 1.813e-05
```

statistically significant predictors in this model. While the coefficient for FTR ($\beta = -1.84$) indicated a potential negative association with retailer counts, the lack of statistical significance suggests that the effect may not be reliably distinguished from zero in this dataset. Similarly, SDI ($\beta = -0.16$) showed no meaningful influence on retailer counts in this analysis. The model as a whole explained a substantial proportion of the variation in retailer counts, with a multiple $R^2 = 0.9652$ and an adjusted $R^2 = 0.9502$, indicating that the included predictors account for about 95% of the variability. The F-statistic ($F = 64.63$, $p < 0.001$) confirmed that the model provides a significant fit to the data, though this may be mainly due to the population input rather than FTRs and SDI scores.

The residuals of this model, shown in the following map, highlight spatial patterns that may indicate the influence of local factors not captured by the model. These patterns can help identify specific municipalities where additional variables or unique local dynamics could be playing a role in retailer distribution, suggesting areas for further investigation and analysis.

OLS Residuals of FTR, SDI, and Population on Brick-and-click Presence Across San Diego County

Effects of flavored tobacco restrictions, social deprivation, and population size on e-commerce tobacco retailer counts.



Grey: No brick-and-click tobacco retailers present.
Date: December 2024

Discussion

This analysis provides valuable insights into the relationship between flavored tobacco restrictions, e-commerce prevalence, and socioeconomic factors. However, several limitations and opportunities for further research have emerged from this study. Future research should incorporate data on "click-only" retailers (tobacco retailers that operate exclusively online with no physical brick-and-mortar establishment), as this analysis focuses solely on brick-and-click operations. Including click-only retailers would offer a more comprehensive understanding of the broader e-commerce landscape and its connection to flavored tobacco restrictions. This would also allow for a better evaluation of the extent to which online sales may circumvent local policies.

In addition, the differences between aggregating the data on the retailer level versus municipality level is something I wish that I had understood better before starting this project. I see now how this can have drastically different implications even when using the same data for each.

I've also realized now that the aggregation of Social Deprivation Index (SDI) scores into averages throughout the municipality level may obscure critical within-municipality socioeconomic disparities. Future analyses should utilize census tract-level SDI data to capture finer-grained variations in socioeconomic deprivation and their potential influence on retailer behaviors. Such an approach would enhance the granularity and relevance of the findings.

While population size emerged as a significant predictor of retailer counts, other potentially influential factors remain unexplored. These include local zoning use types, demographic characteristics such as age distribution and income levels, and the stringency or enforcement of flavored tobacco restrictions. Incorporating these variables in future models could provide a deeper understanding of the mechanisms driving the geographic distribution of e-commerce retailers.

Regulatory loopholes in online sales, particularly those targeting youth populations, warrant immediate attention. Strengthened frameworks should focus on bridging gaps in current policies, with a particular emphasis on jurisdictions with higher youth vulnerability to targeted tobacco marketing. By addressing these gaps, policymakers can better safeguard public health and prevent the exploitation of e-commerce platforms by tobacco retailers.

This study lays a foundation for future exploration into the intersection of public health and tobacco policies, socioeconomics, and e-commerce. Expanding the scope of analysis to other regions, including the extent of all of California, integrating additional variables, and leveraging longitudinal data could yield richer insights and inform more effective regulatory strategies aimed at reducing youth tobacco use and e-commerce exploitation.

References

Bill Text - AB-3218 Unflavored Tobacco List. Accessed December 2024. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB3218#:~:text=This%20bill%20would%20prohibit%20a,not%20listed%20on%20the%20Unflavored

Bill Text - SB-793 Flavored tobacco products. Accessed December 2024. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201920200SB793

California Department of Public Health. (2015). *State health officer's report on e-cigarettes: A community health threat*. California Department of Public Health. https://www.cdph.ca.gov/Programs/CCDPHP/DCDIC/CTCB/CDPH%20Document%20Library/Policy/ElectronicSmokingDevices/StateHealthEcigReport.pdf

California Department of Public Health. (2023). *California youth tobacco survey 2023 annual report*. California Department of Public Health. https://www.cdph.ca.gov/Programs/CCDPHP/DCDIC/CTCB/CDPH%20Document%20Library/ResearchandEvaluation/SurveyInstrumentsTrainingManualsAndProtocols/CYTS2023AnnualReport_FINAL.pdf

Chaffee, B. W., Donaldson, C. D., Couch, E. T., Guerra Castillo, C., Farooq, O., Cheng, N. F., Ameli, N., Wilkinson, M. L., Gansky, S. A., Zhang, X., & others. (2024). Flavored tobacco product use among California adolescents before and immediately after a statewide flavor ban. *Nicotine & Tobacco Research*. https://doi.org/10.1093/ntr/ntae261

Do, E.K., Aarvig, K., Donovan, E.M., Schillo, B.A., Vallone, D.M., Hair E.C. Underage Youth Continue to Obtain E-Cigarettes from Retail Sources in 2022: Evidence from the Truth Continuous Tracking Survey. Int J Environ Res Public Health. 2023 Jan 12;20(2):1399. doi: 10.3390/ijerph20021399. PMID: 36674152; PMCID: PMC9859475.

Gunnell, D., Appleby, L., & Araya, R. (2019). The contribution of social determinants of health to suicide rates: An overview of the evidence. *Social Science & Medicine, 243*, 112551. https://doi.org/10.1016/j.socscimed.2019.112551

Leas, E. C., Mejorado, T., Harati, R., Ellis, S., Satybaldiyeva, N., Morales, N., & Poliak, A. (2023). E-commerce licensing loopholes: A case study of online shopping for tobacco products following a statewide sales restriction on flavoured tobacco in California. *Tobacco Control, 32*(6), 562-568. https://doi.org/10.1136/tc-2023-058269

Tercyak, K. P., Phan, L., Gallegos-Carrillo, K., Mays, D., Audrain-McGovern, J., Rehberg, K., Li, Y., Cartujano-Barrera, F., & Cupertino, A. P. (2020). Prevalence and correlates of lifetime e-cigarette use among adolescents attending public schools in a low-income community in the US. *Addictive Behaviors, 114*, 106738. https://doi.org/10.1016/j.addbeh.2020.106738

The Robert Graham Center, (n.d). *Social deprivation index*. https://www.graham-center.org/maps-data-tools/social-deprivation-index.html

Code with commentary

```r
# Load necessary libraries
library(sf)
library(dplyr)
library(readr)
library(ggplot2)
library(spgwr)
library(sp)
library(spdep)
library(spatialEco)
library(spatstat)

# Load shapefiles
retailers_per_muni <- st_read("~/Documents/GPEC_443/Final/ecomm.shp")
census_sf <- st_read("~/Downloads/calenviroscreen40shpf2021shp/CES4 Final Shapefile.shp")
sdi <- read_csv("~/Documents/GPEC_443/Final/SDI_censustract.csv")
muni_geoms <- st_read("~/Documents/GPEC_443/Final/sd_munis.shp")
population <- read_csv("~/Documents/GPEC_443/Final/Updated_Population_Data.csv")
by_retailer <- st_read("~/Documents/GPEC_443/Final/by_retailer.shp")

# ============================================================================
# Data Preparation, Aggregated by Muni and Aggregated by Retailer (by_retailer)
# ============================================================================

# Clean and prepare retailers_per_muni
retailers_per_muni <- retailers_per_muni %>%
  mutate(name = ifelse(name == "S.D. COUNTY", "Unincorporated Areas", name)) %>%
  mutate(name = tools::toTitleCase(tolower(name))) %>%
  st_drop_geometry()

# Filter census data for San Diego County
census_sf_sd <- census_sf %>%
  filter(County == "San Diego") %>%
  st_drop_geometry()  # Drop geometry as it's not needed

# Prepare SDI data
sdi <- sdi %>%
  rename(Tract = CENSUSTRACT_FIPS)  # Rename for joining

# Merge census data with SDI data
sdi_tracts <- census_sf_sd %>%
  left_join(sdi, by = "Tract") %>%
  dplyr::select(ApproxLoc, SDI_score) %>%
  mutate(SDI_score = as.numeric(SDI_score))

# Aggregate SDI scores by municipality
sdi_aggregated <- sdi_tracts %>%
  group_by(ApproxLoc) %>%
  summarize(average_sdi_score = mean(SDI_score, na.rm = TRUE))

# Match SDI scores to retailers_per_muni
retailers_per_muni <- retailers_per_muni %>%
  left_join(sdi_aggregated, by = c("name" = "ApproxLoc"))

# Handle unmatched rows by adding them to Unincorporated Areas
unmatched_rows <- sdi_aggregated %>%
  filter(!ApproxLoc %in% retailers_per_muni$name)
```

```r
unincorporated_row <- retailers_per_muni %>%
  filter(name == "Unincorporated Areas") %>%
  mutate(average_sdi_score = mean(unmatched_rows$average_sdi_score, na.rm = TRUE))

retailers_per_muni <- retailers_per_muni %>%
  filter(name != "Unincorporated Areas") %>%
  bind_rows(unincorporated_row)

# Add population data to retailers_per_muni
retailers_per_muni <- retailers_per_muni %>%
  left_join(population, by = c("name" = "Municipality"))

# Calculate retailer rate (retailers per 100,000 population)
retailers_per_muni <- retailers_per_muni %>%
  mutate(retailers_per_100k = (retailers / Population) * 100000)

# Log-transform retailer rate
retailers_per_muni <- retailers_per_muni %>%
  mutate(log_retailers_per_100k = ifelse(
    retailers_per_100k == 0, NA, log(retailers_per_100k)
  ))

# Calculate the SDI indice (normalized SDI score)
retailers_per_muni <- retailers_per_muni %>%
  mutate(
    sdi_indice = (average_sdi_score - mean(average_sdi_score, na.rm = TRUE)) /
      sd(average_sdi_score, na.rm = TRUE)
  )

# Aggregate geometries by municipality
aggregated_geoms <- muni_geoms %>%
  group_by(name) %>%
  summarize(geometry = st_union(geometry))

# Clean `name` column in `retailers_per_muni`
retailers_per_muni <- retailers_per_muni %>%
  mutate(name = toupper(name)) %>%
  mutate(name = ifelse(name == "UNINCORPORATED AREAS", "S.D. COUNTY", name))

# Perform the left join
muni_data <- aggregated_geoms %>%
  left_join(retailers_per_muni, by = "name")

# Filter for complete cases
analyze <- muni_data %>%
  filter(!is.na(log_retailers_per_100k) & !is.na(sdi_indice))

# Add the log_retailers column to analyze
analyze <- analyze %>%
  mutate(log_retailers = ifelse(retailers == 0, NA, log(retailers)))

# Merge on "name"
analyze_ret <- by_retailer %>%
  left_join(aggregated_geoms %>% st_drop_geometry(), by = "name")

# Add the polygon geometry from aggregated_geoms to the merged table
analyze_ret <- analyze_ret %>%
  mutate(geometry = st_geometry(aggregated_geoms)[match(name, aggregated_geoms$name)])
```

```r
# Ensure the resulting table is an sf object
analyze_ret <- st_as_sf(analyze_ret)

muni_data <- muni_data %>%
  mutate(retailers_z_score = (retailers - mean(retailers, na.rm = TRUE)) /
           sd(retailers, na.rm = TRUE))

# Ensure both are in the same CRS (Coordinate Reference System)
muni_data <- st_transform(muni_data, crs = st_crs(by_retailer))

# Perform the spatial join
final_table <- st_join(by_retailer, muni_data, left = TRUE)

# Select the desired columns
final_table <- final_table %>%
  select(lat, lon, name.x, FTR.x, average_sdi_score, Population, sdi_indice, retailers_z_score)

# saving a sf
#st_write(muni_data, "~/Documents/GPEC_443/Final/muni_sdi.shp")

# =========================
# Welch's Two Sample T-test
# =========================

t_test_z <- t.test(retailers_per_100k ~ FTR, data = muni_data)
print(t_test_z)

# =================
# Nearest Neighbors
# =================

st_crs(by_retailer)

# Extract coordinates
coords <- st_coordinates(final_table)

# Define the bounding window
window <- owin(
  xrange = range(coords[, 1]),
  yrange = range(coords[, 2])
)

# Convert to ppp object
retailer_ppp <- as.ppp(coords, W = window)

# Nearest neighbor distances
nn_dist <- nndist(retailer_ppp)
summary(nn_dist)

# Calculate observed mean nearest neighbor distance
observed_mean_distance <- mean(nn_dist)

# Calculate expected mean nearest neighbor distance for a Poisson process
area <- area.owin(retailer_ppp$window)
expected_mean_distance <- 1 / (2 * sqrt(npoints(retailer_ppp) / area))

# Compute Nearest Neighbor Index (NNI)
nni <- observed_mean_distance / expected_mean_distance
```

```
# Interpret the NNI
if (nni < 1) {
  interpretation <- "Clustered pattern"
} else if (nni == 1) {
  interpretation <- "Random distribution"
} else {
  interpretation <- "Regular pattern"
}

# Print results
cat("Observed Mean Distance:", observed_mean_distance, "\n")
cat("Expected Mean Distance:", expected_mean_distance, "\n")
cat("Nearest Neighbor Index (NNI):", nni, "\n")
cat("Interpretation:", interpretation, "\n")

# ====================
# OLS Regression, Muni
# ====================

# Perform the OLS regression
ols_model <- lm(retailers ~ FTR + sdi_indice + Population, data = analyze)

# Print the summary of the regression model
summary(ols_model)

# Calculate residuals from ols_model2 and add them to a new data frame
residuals_df <- analyze %>%
  mutate(ols_residuals = residuals(ols_model)) %>%  # Reference the correct model
  select(name, ols_residuals) %>%  # Keep only name and residuals for joining
  st_drop_geometry() %>%           # Drop geometry to avoid conflicts
  as.data.frame()                  # Convert to a standard data frame

# Merge the residuals back into muni_data using st_join
muni_data <- muni_data %>%
  left_join(residuals_df, by = "name")

# Create the map
ols_plot <- ggplot() +
  geom_sf(
    data = muni_data %>% filter(!is.na(ols_residuals)),
    aes(fill = ols_residuals),
    color = "grey44",
    size = 0.0005
  ) +
  scale_fill_distiller(
    palette = "RdBu",
    limits = c(-1.8, 1.8),
    breaks = seq(-1.8, 1.8, by = 0.9),
    name = "OLS Residuals\n(Retailer Presence vs\nFTR, SDI, & Population)"
  ) +
  geom_sf(
    data = muni_data %>% filter(is.na(ols_residuals)),
    fill = "grey",
    color = "gray44",
    size = 0.0005
  ) +
```

```r
  labs(
    title = "OLS Residuals of FTR, SDI, and Population on Brick-and-click Presence Across San Diego County",
    subtitle = "Effects of flavored tobacco restrictions, social deprivation, and population size on e-commerce tobacco retailer counts.",
    caption = "Grey: No brick-and-click tobacco retailers present.\nDate: December 2024"
  ) +
  annotation_north_arrow(
    location = "tr",
    which_north = "true",
    style = north_arrow_fancy_orienteering()
  ) +
  annotation_scale(
    location = "br",
    width_hint = 0.25
  ) +
  theme_minimal() +
  theme(
    legend.position = c(0.12, 0.18),
    legend.background = element_rect(fill = alpha("white", 0.8), color = NA),
    legend.key.size = unit(0.8, "lines"),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
    plot.title = element_text(size = 14, face = "bold"),
    plot.subtitle = element_text(size = 11),
    panel.grid = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    panel.border = element_rect(color = "black", fill = NA, size = 2)
  )

# Print and save the plot
print(ols_plot)
ggsave("~/Documents/GPEC_443/Final/ols_residuals_map_2.png", plot = ols_plot, width = 10, height = 8, dpi = 300)

# ==============
# Kernal density
# ==============

# Reproject spatial data to UTM Zone 11N
final_table <- st_transform(final_table, crs = 32611)

# Reproject muni_data to UTM Zone 11N
muni_data <- st_transform(muni_data, crs = 32611)

# Create unique points dataset by removing duplicates
final_table_unique <- final_table %>%
  distinct(geometry, .keep_all = TRUE)

# Step 1: Define the Extent from muni_data
# Get bounding box from muni_data
muni_bbox <- st_bbox(muni_data)
```

```
# Create a window from the bounding box
window <- owin(
  xrange = c(muni_bbox["xmin"], muni_bbox["xmax"]),
  yrange = c(muni_bbox["ymin"], muni_bbox["ymax"])
)

# Convert retailer points to ppp format
coords <- st_coordinates(final_table_unique)  # Use your unique points dataset
ppp_data <- ppp(x = coords[, 1], y = coords[, 2], window = window)

# Perform KDE
kde <- density(ppp_data, sigma = 1421, edge = TRUE)  # Adjust sigma as needed

# Convert KDE to raster
kde_raster <- raster(kde)

# Update CRS of the raster to match muni_data
crs(kde_raster) <- st_crs(muni_data)$proj4string

# Convert density to per 10 square kilometers
kde_raster_10km2 <- kde_raster * 1e7

# Debugging Step: Plot KDE raster before masking
kde_raster_df <- as.data.frame(kde_raster_10km2, xy = TRUE)
kde_raster_df <- kde_raster_df %>% na.omit()
colnames(kde_raster_df) <- c("x", "y", "density")

initial_plot <- ggplot() +
  geom_raster(data = kde_raster_df, aes(x = x, y = y, fill = density)) +
  geom_sf(data = muni_data, fill = NA, color = "black", size = 0.5) +
  scale_fill_viridis_c(option = "D", name = "Density (per 10 km²)", direction = -1) +
  labs(
    title = "Kernel Density of Retailers (Before Masking)",
    x = "Longitude", y = "Latitude"
  ) +
  theme_minimal()

print(initial_plot)

# Mask Raster to Municipal Boundaries
muni_sp <- as(st_geometry(muni_data), "Spatial")  # Convert sf to Spatial object
kde_raster_masked <- mask(kde_raster_10km2, muni_sp)

# Convert masked raster to a data frame for ggplot2
kde_df_masked <- as.data.frame(kde_raster_masked, xy = TRUE)
kde_df_masked <- kde_df_masked %>% na.omit()
colnames(kde_df_masked) <- c("x", "y", "density")
```

```r
# Plot the KDE with Municipal Boundaries
kde_plot <- ggplot() +
  # Raster layer (KDE)
  geom_raster(data = kde_df_masked, aes(x = x, y = y, fill = density)) +
  # Municipal boundaries
  geom_sf(data = muni_data, fill = NA, color = "black", size = 0.5) +
  # Adjust color scale
  scale_fill_viridis_c(option = "D", name = "Density (per 10 km²)", direction = -1) +
  # Add labels and theme
  labs(
    title = "Kernel Density of Brick-and-click Retailers (Masked to Municipal Boundaries)",
    x = "Longitude", y = "Latitude"
  ) +
  theme_minimal()

print(kde_plot)
# Save the plot
ggsave("kde_density_plot.png", plot = kde_plot, width = 10, height = 8, dpi = 300)

# =========================================================================
# Global Moran's I, Muni, test if retailer rates exhibit spatial clustering
# =========================================================================

# Create a new spatial weights matrix based on the filtered dataset
nb_analyze <- poly2nb(analyze, queen = TRUE)
listw_non_na <- nb2listw(nb_analyze, style = "W")

# Moran's I test
moran_test <- moran.test(analyze$retailers_per_100k, listw_non_na)

# Print the result
print(moran_test)

# Visualize spatial autocorrelation
library(spatialEco)
moran.plot(
  analyze$retailers_per_100k,
  listw_non_na,
  labels = analyze$name,
  xlab = "Retailer Rates",
  ylab = "Spatially Lagged Rates",
  main = "Moran's I Plot for Retailer Rates"
)
```