Final Paper

for INF 412 (J. Wang)
March 26, 2023

Nicholas Bosco, Oliver Daniel, Tiago Martins, Yahui Zhang

Introduction

With currently 417 marijuana dispensaries in the city of Toronto, and more likely to be opened in the near future, it begs the question as to where the next dispensary might open. The Alcohol and Gaming Commission of Ontario's iAGCO map displays currently opened dispensaries, and while this information is important to consider, it offers little in the way of ascertaining where a new dispensary might open next. Our group is curious as to what factors influence the location of a new dispensary, particularly the presence of different types of businesses in the area. By performing a logistic regression, our group hopes to predict the likeliest location for a new dispensary based on different types of nearby businesses.

Background

First among the two sources we consulted for this assignment was Planning for Marijuana: The Cannabis Conundrum, by Jeremy Németh & Eric Ross. The stated purpose of this article was to explore the local regulation of medical marijuana dispensaries in various states in the US, and by what factors are different zones determined as eligible for new dispensaries, and which factors inhibit the construction of a medical marijuana dispensary (MMD). Of information that was particularly relevant to use, this report found that so-called "buffers" severely diminished an MMD's chances of being built in a given neighborhood. These buffers included the existence of other MMDs, but also "sensitive facilities" such as childcare centres, churches, libraries, parks, and also cinemas, recreation centres, and tobacco stores. Furthermore, these buffer businesses included bars, fast food restaurants, and liquor stores on the basis of resident's concerns of crime and further diminishing property values.

Our second source is Locating Medical and Recreational Cannabis Outlets for Research Purposes: Online Methods and Observational Study, an effort to determine the means of locating and observing both licensed and unlicensed dispensaries in Los Angeles County. While the main objective of this study was to determine which of the located dispensaries were open for business, the study also contained additional details regarding the dispensaries in question. Of interest to us was the finding of how 40.6% of identified dispensaries had a tobacco store in close proximity.

Data

Dispensaries

In the province of Ontario, all cannabis retail locations must be issued a license by the Alcohol and Gaming Commission of Ontario (AGCO) before opening. If the application is not rejected outright, the location owners must place a 'Public Notice' placard in their proposed store for a 15-day period before the ACGO will consider approving the application and allowing cannabis sales to begin. The ACGO maintains an online

directory of all current applications, regardless of status, across the province, as well as a corresponding interactive map.

This map, ulimately, is fed by a CSV file on the ACGO webserver. The URL to this CSV is accessible to the public without credentials, and is available for download; see line 11 in `scripts/fetch-and-clean-data.r`. Within is a wealth of data for Ontario dispensary applications, including names, addresses, application status, geographic coordinates, and more. For the purposes of this paper, we excluded any application which did not satisfy the following:

1. The application status must be **Authorized to Open**, i.e., neither rejected nor currently in public notice at the time of download; and
2. The dispensary's postal code must begin with **M**, which indicates that the address is serviced by a Toronto FSA[1].

Filtering by these two criteria, we were left with a total of 416 open dispensaries within the City of Toronto, from each of which we retained only the FSA and, for plotting purposes, latitude and longitude. Other identifying details, such as specific addresses and store names, were discarded.

Businesses

Through the Open Data Toronto initiative, the City of Toronto offers to the public a listing of all current business licenses[2] issued within the Greater Toronto Area. Along with business information and addresses, the data also featured the `category` of business under which the issued license falls. These are also available for download in multiple useful formats, such as CSV.

The data were filtered again by the first letter of their postal code to limit addresses to those within the City of Toronto. However, this still resulted in 10 different business categories. For the purpose of this paper, we limited our data to only 10:

name
HOLISTIC CENTRE
EATING ESTABLISHMENT
ENTERTAINMENT ESTABLISHMENT/NIGHTCLUB
PAWN SHOP
PAYDAY LOAN
SIDEWALK CAFE
RETAIL STORE (FOOD)
SMOKE SHOP
VAPOUR PRODUCT RETAILER
ADULT ENTERTAINMENT CLUB

For a total of 69,141 suitable businesses. For each of these businesses, only the FSA and category were retained.

Business types   These ten business types were chosen through a combination of our background research, such as including vapour products in relation to Pedersen et al. (2020); total quantities throughout Toronto to prevent data scarcity issues with prediction; and intuition. [TODO more]

---

[1]This is a useful heuristic for discriminating what addresses are "within" the City of Toronto.
[2]i.e., those private business licenses that can be issued at the City's discretion, excluding e.g., liquor and cannabis retail locations.

Forward Sortation Areas (FSAs)

Using data drawn from the Census, the Canadian federal government subdivides the geography of Canada into 1,620 unique **Forward Sortation Areas**, or FSAs. Each address found within an FSA shares the same first three characters of their postal code, which makes these three characters a convenient shorthand for a particular region of Canada for purposes of shipping, travel, and more. Conveniently, the City of Toronto is so large that it gets its own prefix for FSAs: M. This is why the dispensary and business license data were filtered to those whose postal codes began with M above.

For the purposes of this paper, a series of shapefiles (DMTA, Inc., 2005) outlining the FSAs was downloaded from the UofT digital library. Although the file was uploaded in 2005 and Toronto's FSAs have changed slightly in 2023, we considered this a minor issue for the purposes of our experiment. Note that, although the source code[3] for this paper is available online, this shapefile is proprietary and can only be accessed by UofT faculty and staff with valid credentials.

Methodology

Results

Scatter plots

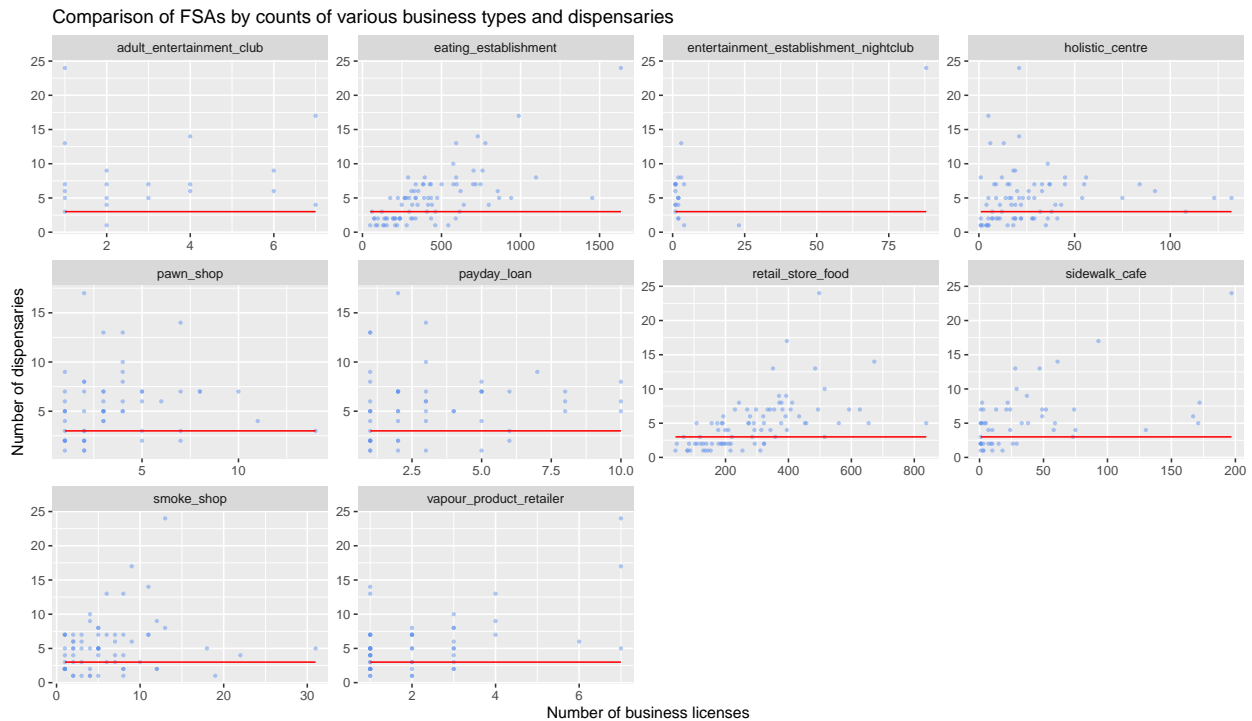For the figures shown below in Figure @ref(fig:scatterplots), note the free scales for both axes.

Figure 1: The red horizontal line represents the delineation between class-0 and class-1 FSAs: the dots found above the line represent those FSAs which have at least 3 dispensaries, and vice versa.
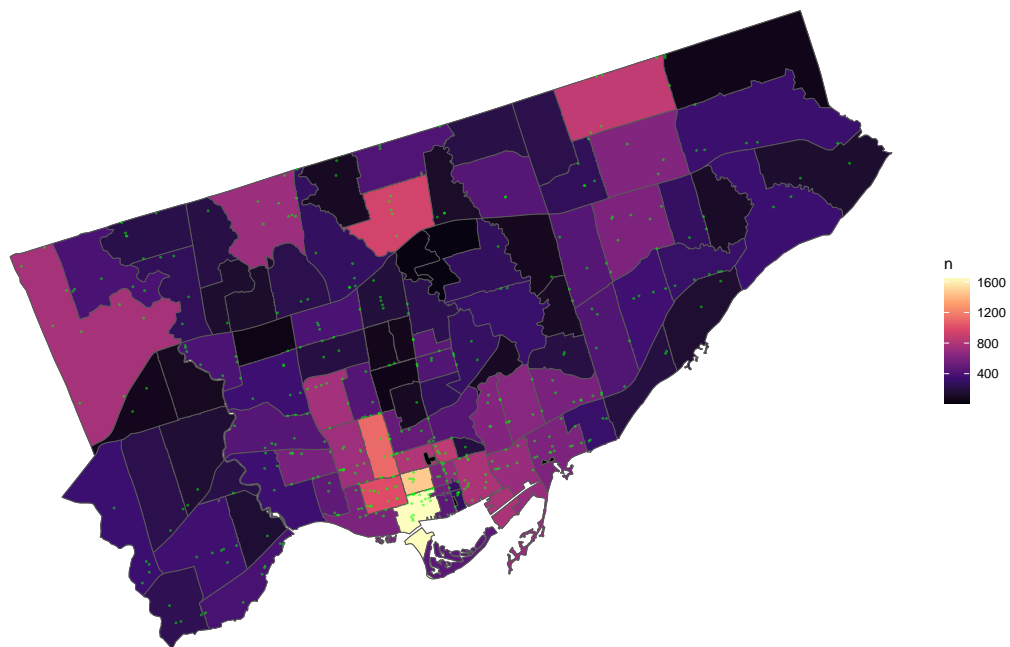
---

[3]https://github.com/oliver-daniel/inf_412_final_project
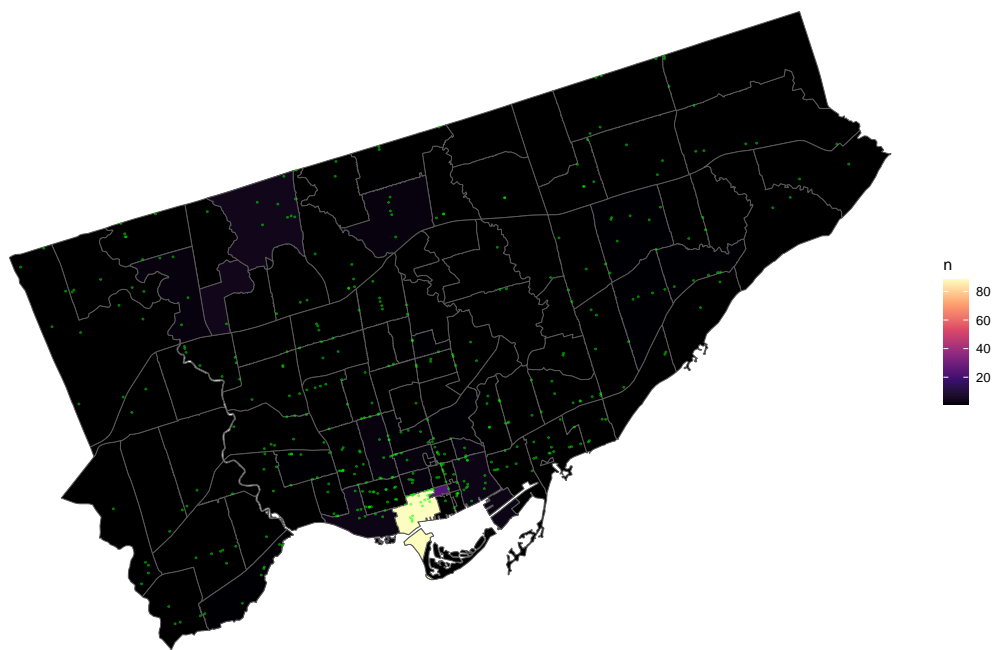
Spatial Visualization

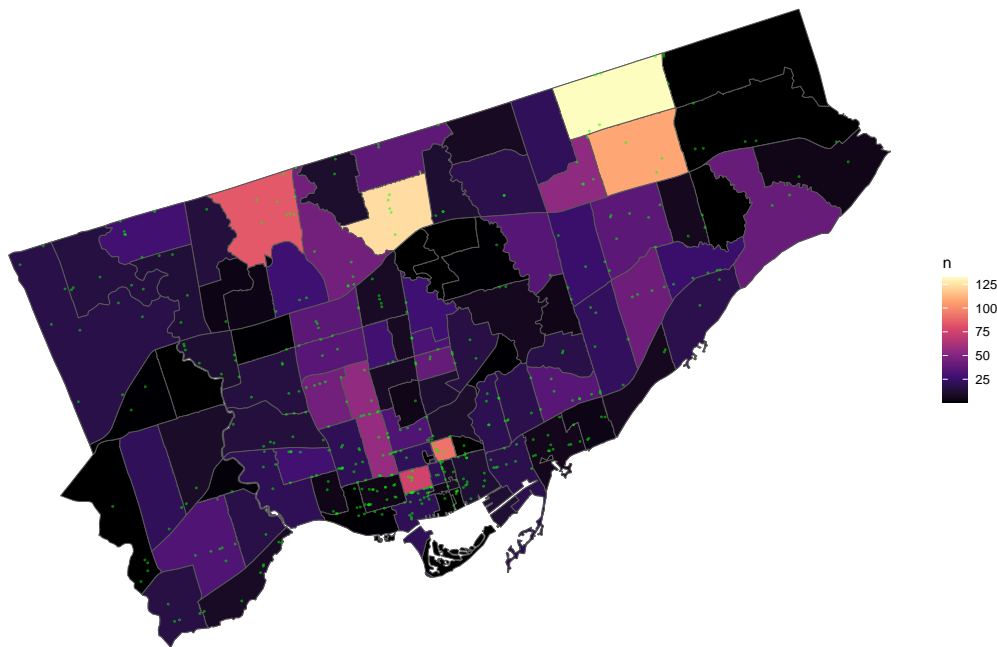Count map of Adult entertainment club per FSA of Toronto
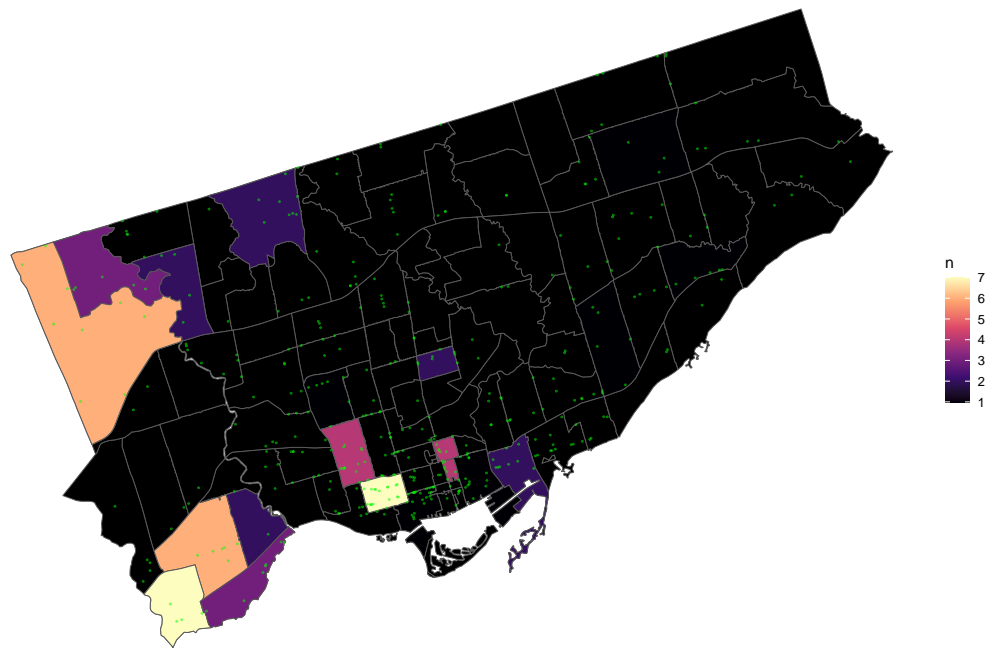
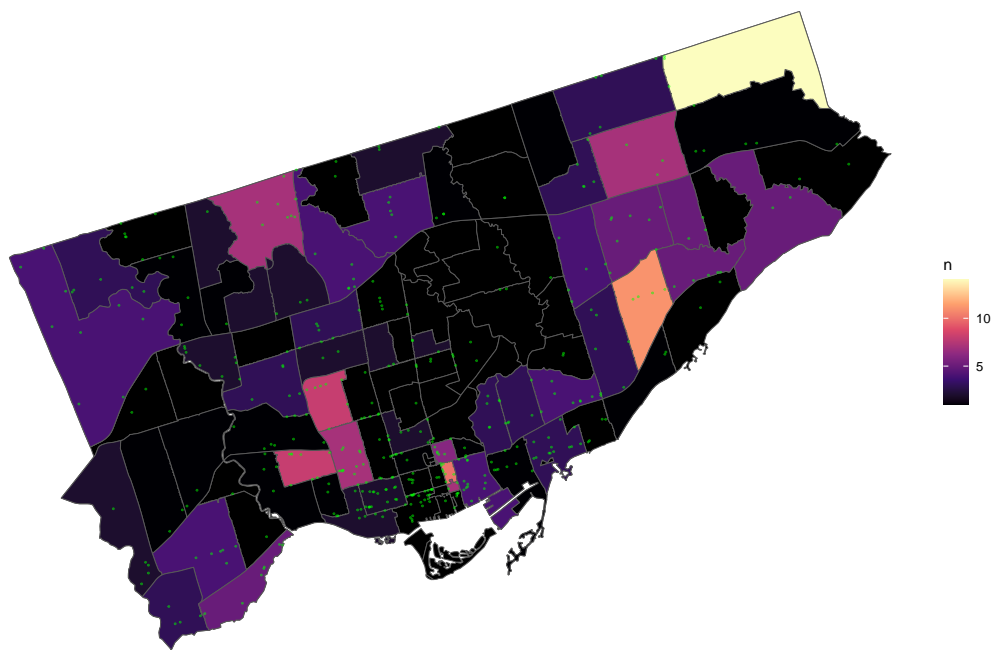Count map of Eating establishment per FSA of Toronto

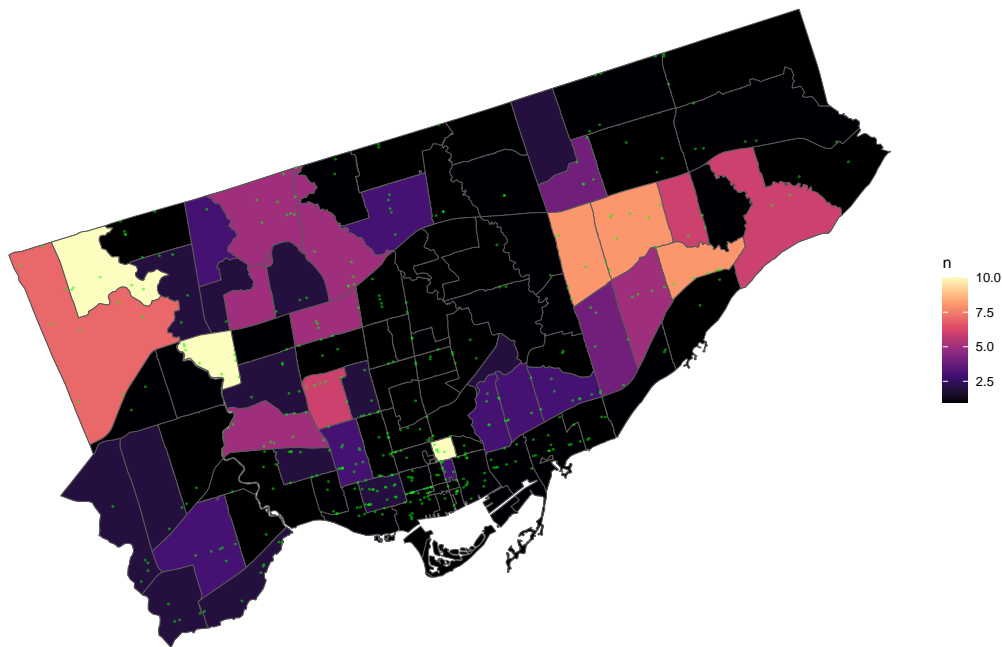Count map of Entertainment establishment nightclub per FSA of Toronto

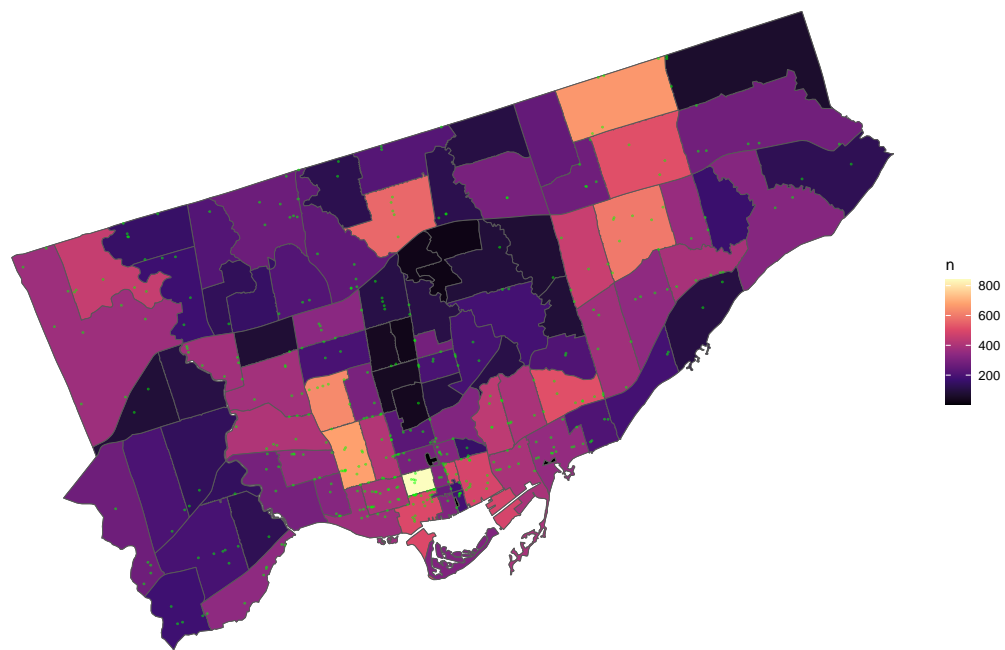Count map of Holistic centre per FSA of Toronto
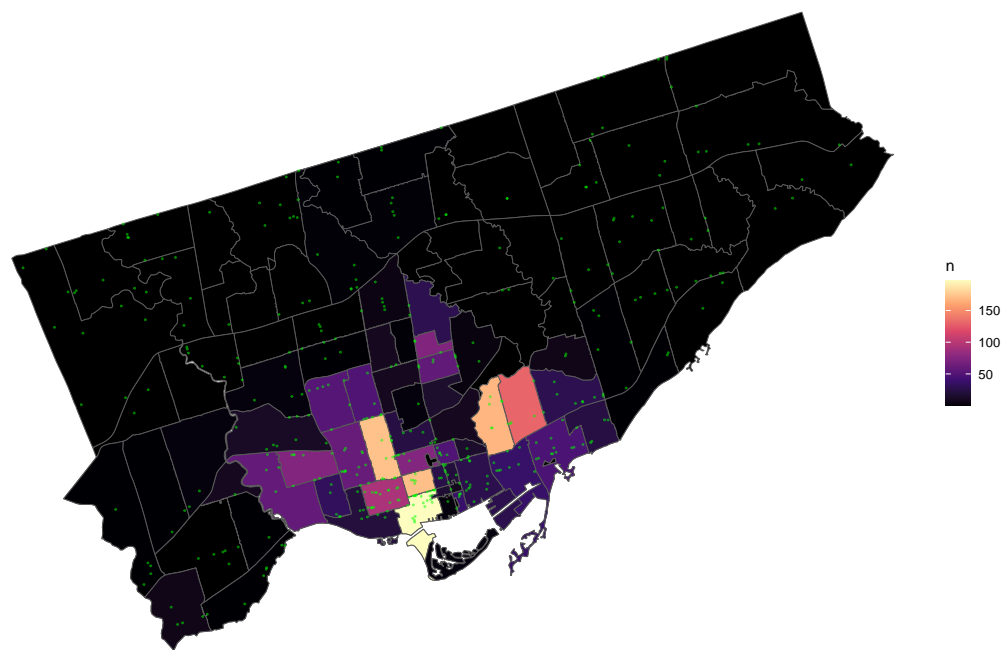
Count map of Pawn shop per FSA of Toronto

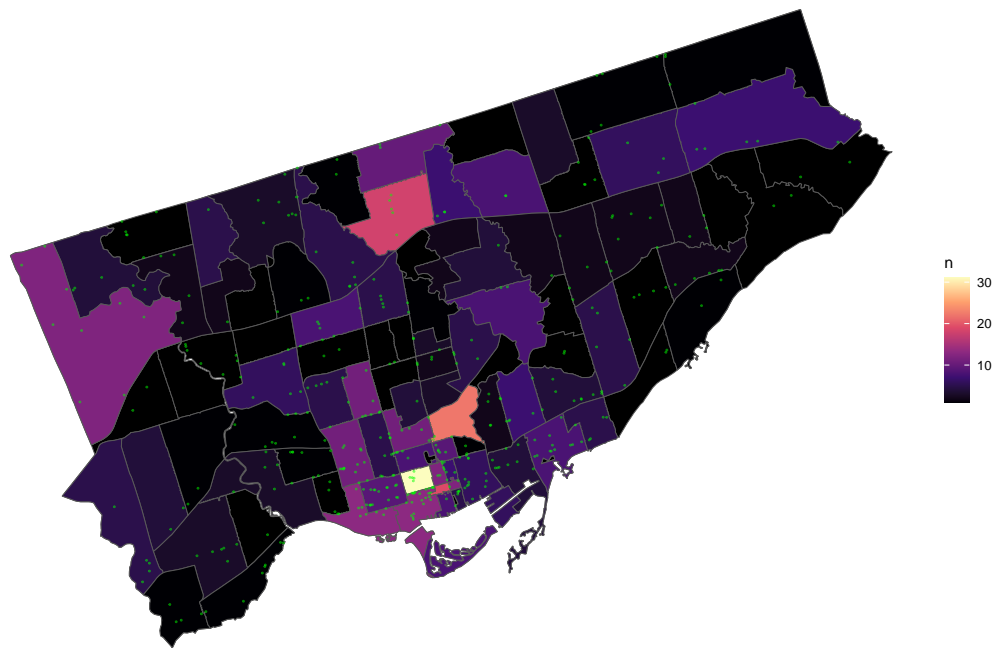Count map of Payday loan per FSA of Toronto
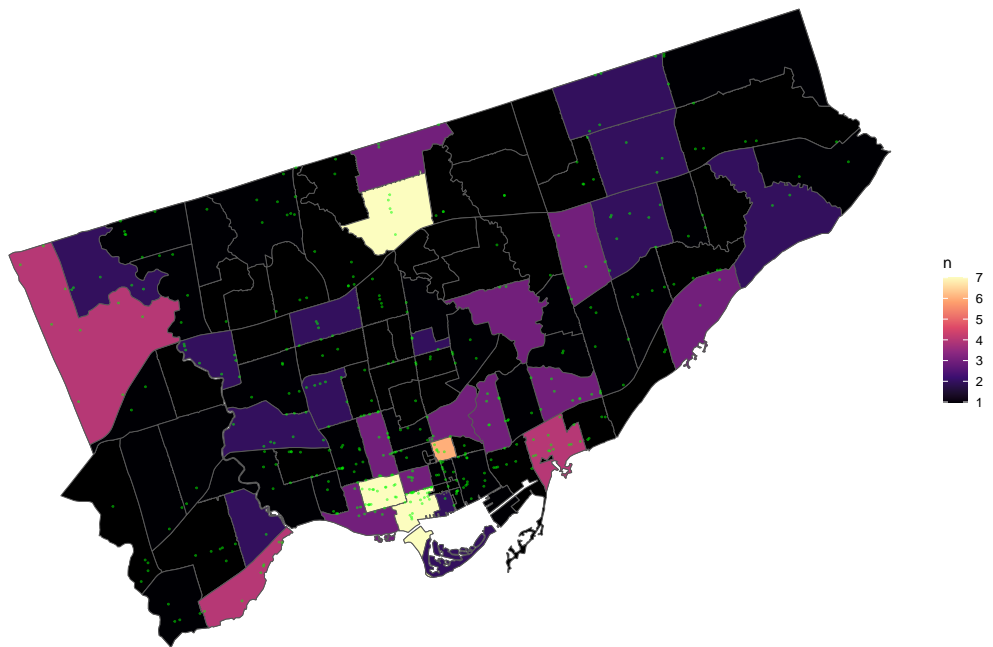
Count map of Retail store food per FSA of Toronto

Count map of Sidewalk cafe per FSA of Toronto

Count map of Smoke shop per FSA of Toronto

Count map of Vapour product retailer per FSA of Toronto

Data Summary

Table 2: Data Summary

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
n.dispensaries	15	0	4.1	3.9	0.0	3.0	24.0	
class	2	0	0.5	0.5	0.0	1.0	1.0	
eating_establishment	94	0	371.1	297.5	0.0	308.5	1635.0	
holistic_centre	45	0	21.1	25.3	0.0	15.0	132.0	
pawn_shop	12	0	2.0	2.7	0.0	1.0	14.0	
payday_loan	10	0	1.8	2.5	0.0	1.0	10.0	
retail_store_food	94	0	254.4	167.8	0.0	218.5	838.0	
smoke_shop	18	0	4.4	5.2	0.0	3.0	31.0	
vapour_product_retailer	7	0	1.2	1.6	0.0	1.0	7.0	
adult_entertainment_club	7	0	0.6	1.5	0.0	0.0	7.0	
entertainment_establishment_nightclub	7	0	1.4	9.0	0.0	0.0	88.0	
sidewalk_cafe	37	0	19.6	39.7	0.0	1.0	197.0	

Logistical Model

Logistical Model before AIC

```
##
## Call:
## glm(formula = class ~ adult_entertainment_club + eating_establishment +
##     entertainment_establishment_nightclub + pawn_shop + payday_loan +
##     sidewalk_cafe + retail_store_food + smoke_shop + vapour_product_retailer +
##     holistic_centre, family = binomial, data = fsa_counts_pivot)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.17463  -0.36213   0.00316   0.27078   2.27962
##
## Coefficients:
##                                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)                            -4.327123   0.989693  -4.372 1.23e-05 ***
## adult_entertainment_club                1.336171   0.766515   1.743   0.0813 .
## eating_establishment                   -0.003192   0.007176  -0.445   0.6565
## entertainment_establishment_nightclub  -0.151355   0.088613  -1.708   0.0876 .
## pawn_shop                               0.307465   0.142403   2.159   0.0308 *
## payday_loan                             0.084086   0.252652   0.333   0.7393
## sidewalk_cafe                           0.079679   0.039893   1.997   0.0458 *
## retail_store_food                       0.013875   0.009294   1.493   0.1355
## smoke_shop                              0.137796   0.125182   1.101   0.2710
## vapour_product_retailer                -0.005302   0.402971  -0.013   0.9895
## holistic_centre                         0.012478   0.025183   0.495   0.6203
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 140.420  on 101  degrees of freedom
## Residual deviance:  54.726  on  91  degrees of freedom
## AIC: 76.726
##
## Number of Fisher Scoring iterations: 7
```

Logistical Model after AIC

```
##
## Call:
## glm(formula = class ~ adult_entertainment_club + entertainment_establishment_nightclub +
##     pawn_shop + sidewalk_cafe + retail_store_food + smoke_shop,
##     family = binomial, data = fsa_counts_pivot)
##
## Deviance Residuals:
##     Min        1Q    Median        3Q       Max
## -2.24204  -0.35733   0.00313   0.24148   2.21545
##
## Coefficients:
##                                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)                          -4.343323   0.940682  -4.617 3.89e-06 ***
## adult_entertainment_club              1.260875   0.716272   1.760  0.07835 .
## entertainment_establishment_nightclub -0.158938   0.082962  -1.916  0.05539 .
## pawn_shop                             0.333935   0.145720   2.292  0.02193 *
## sidewalk_cafe                         0.065309   0.030859   2.116  0.03431 *
## retail_store_food                     0.012686   0.003958   3.205  0.00135 **
## smoke_shop                            0.081374   0.086408   0.942  0.34632
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 140.420  on 101  degrees of freedom
## Residual deviance:  55.291  on  95  degrees of freedom
## AIC: 69.291
##
## Number of Fisher Scoring iterations: 7
```

Exponentiate the coefficients

```
##                           (Intercept)          adult_entertainment_club
##                            0.01299327                        3.52850902
## entertainment_establishment_nightclub                         pawn_shop
##                            0.85304940                        1.39645304
##                         sidewalk_cafe                 retail_store_food
##                            1.06748835                        1.01276725
##                            smoke_shop
##                            1.08477647
```

K-fold cross-validation

```
## Generalized Linear Model
##
## 102 samples
##   6 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 91, 91, 93, 93, 92, 92, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.8521212  0.7068956
```

Discussion

Conclusion

Our regression model indicates that at least certain types of businesses are significantly ($p < 0.05$) correlated to the likelihood of dispensaries being opened in an arbitrary FSA-sized region of Toronto. [todo more about AIC, methodologies, etc.] Where higher numbers of dispensaries have already been opened, often the most densely populated regions of the city, co-location of multiple dispensaries in close proximity may be indicative of the longer-term stability of retail cannabis in those areas. However, statistical modelling can prove to be an effective tool in determining which geographic regions may be hospitable to a burgeoning cannabis market, inviting early investment and temporarily enjoying the dividends of low competition.

Considerations

Data is sparse

Being only a few years removed from legalization, cannabis retail is still in its nascent stages, with even more veteran markets being only a few years its senior. The scene is changing daily, and it is difficult even for licensing bodies like AGCO to track the opening and closing of new locations. This is especially true of grey-market dispensaries, such as First Nations-owned locations which operate on the fringes of treaty law. Although efforts exist to collect and list them, primarily for consumer-facing purposes, there is a distinct lack of verified, up-to-date data on these locations. It would be of particular interest to investigate these locations and their market penetration through further study.

Possible confounding factors

The scope of our study fails to control for a number of important factors, which could bear significant impact on our statistical findings. These include:

Population The high variance in population density across the City of Toronto is a latent variable that affects the density of dispensaries and other businesses alike: where there are more people, there will certainly be more establishments to serve them. It is unknown to what extent our model accounts for this, and it is possible that certain signficant factors for dispensary presence in an FSA could be reliant on data only from downtown regions.

Regulations and public infrastructure Our dataset includes only private business addresses, excluding non-business institutions such as schools, places of worship, and entrances to public and government infrastructure. Future research into the legal constraints of dispensary location – distance from schools, by-law regulations, etc. – may indicate other forces acting on the location of the next dispensary when estimating.

Geospatial granularity The choice to use the 102 FSAs[4] of the City of Toronto as geographic subdivisions was a largely arbitrary one, looking to balance predictive power with visibility and performance when authoring this paper. Other options considered included the electoral ridings or neighbourhood designations of the city; an arbitrary quadrilateral or hexagonal grid, overlain atop a map of Toronto; and even a Voronoi cell system connecting adjacent street intersections into "blocks". Each of these was deemed unsuitable for varied reasons, but future reproductive studies may wish to test whether our results are repeatable on coarser- or finer-grained geographic scales.

---

[4]As of 2005.

Pedersen, Eric R, Caislin Firth, Jennifer Parker, Regina A Shih, Steven Davenport, Anthony Rodriguez, Michael S Dunbar, et al. 2020. "Locating Medical and Recreational Cannabis Outlets for Research Purposes: Online Methods and Observational Study." *Journal of Medical Internet Research* 22 (2). https://doi.org/10.2196/16853.