Final Paper

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

Nicholas Bosco, Oliver Daniel, Tiago Martins, Yahui Zhang

Contents

Introduction

Methodology

Results

Scatter plots

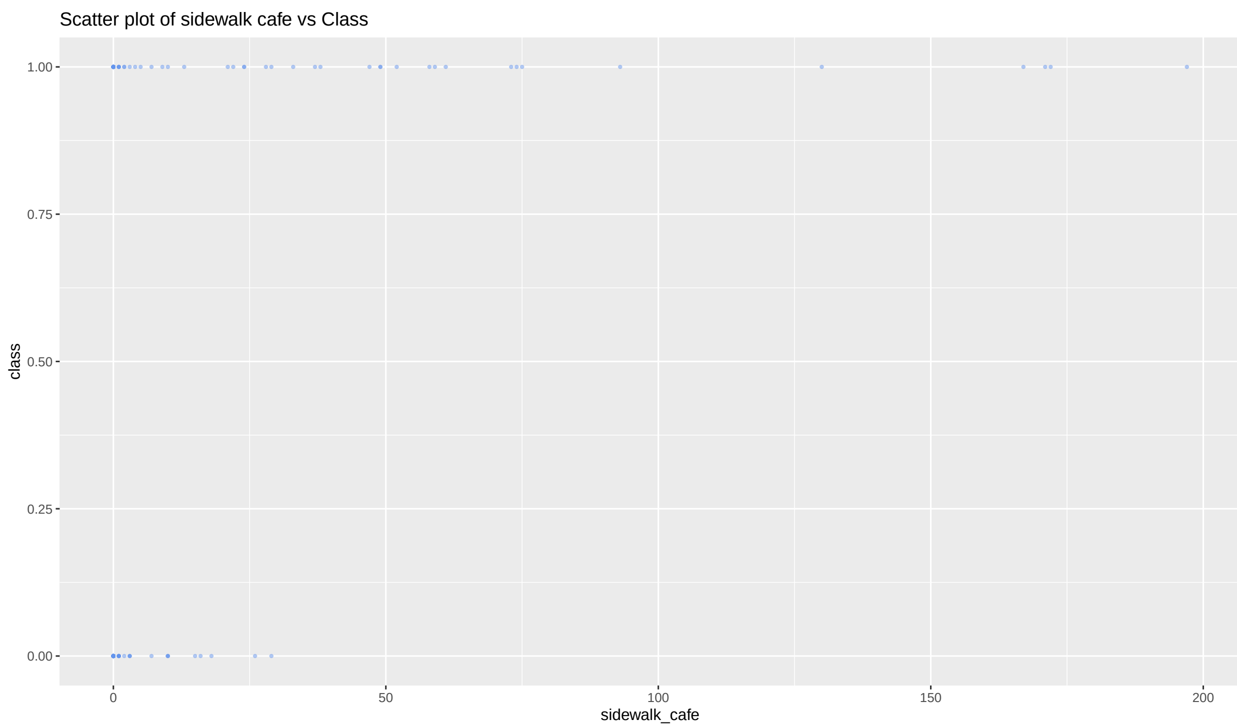


Figure 1: Sidewalk cafe vs dispensaries

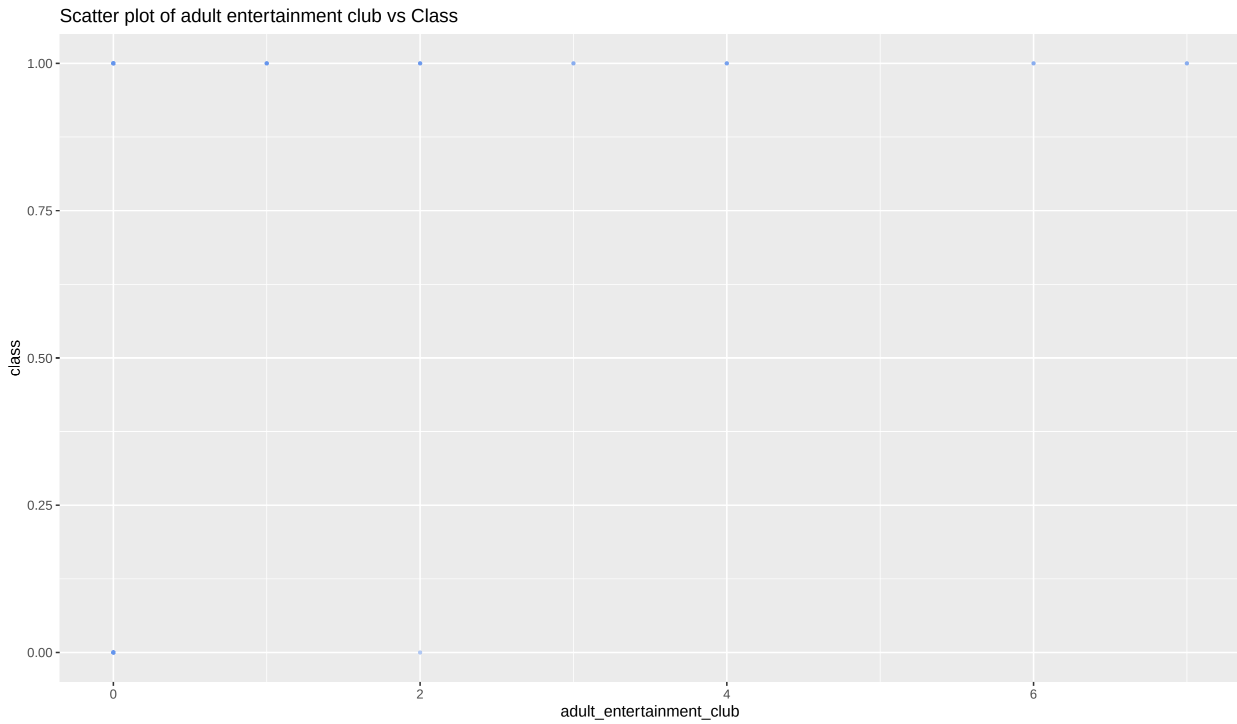


Figure 2: Adult entertainment club vs dispensaries

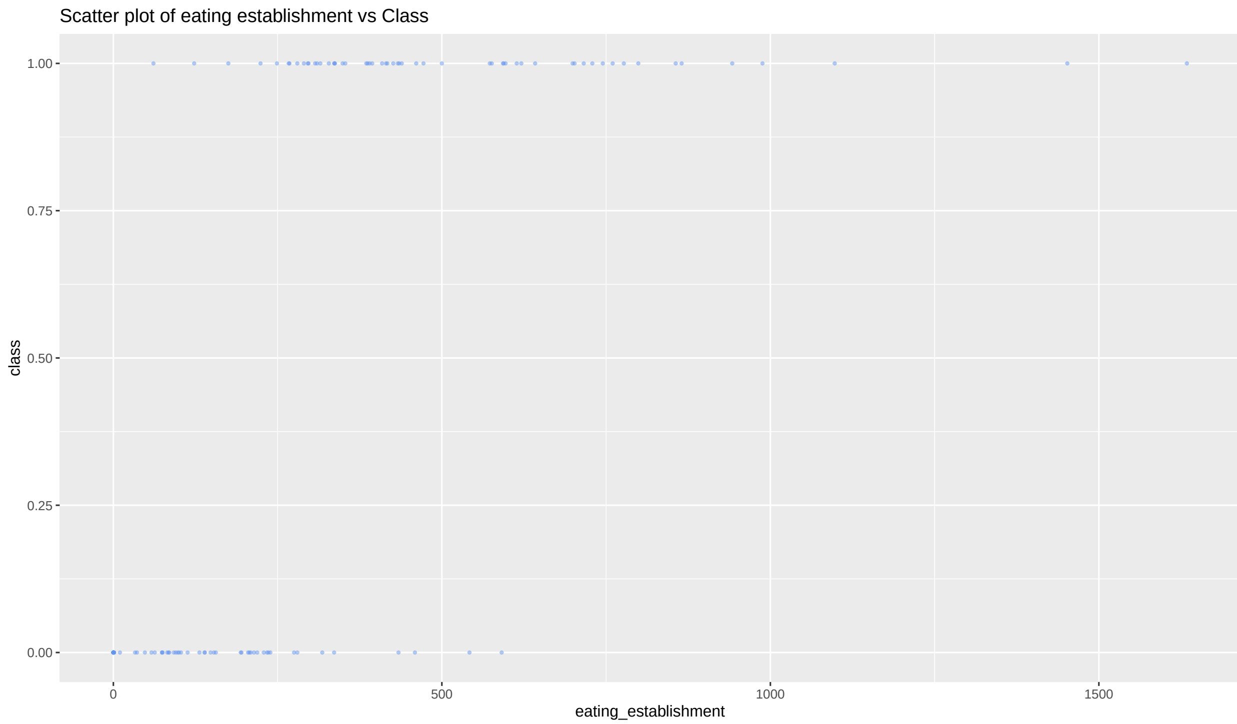


Figure 3: Eating establishment vs dispensaries

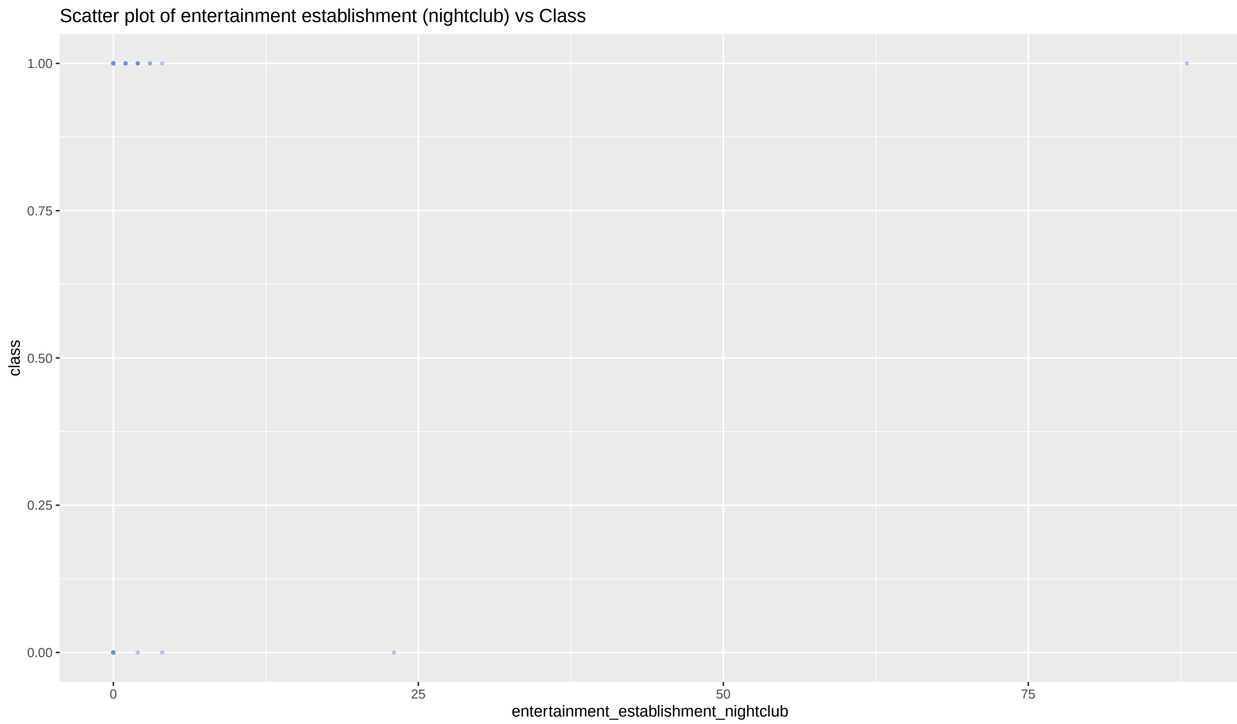


Figure 4: Entertainment establishment (nightclub) vs dispensaries

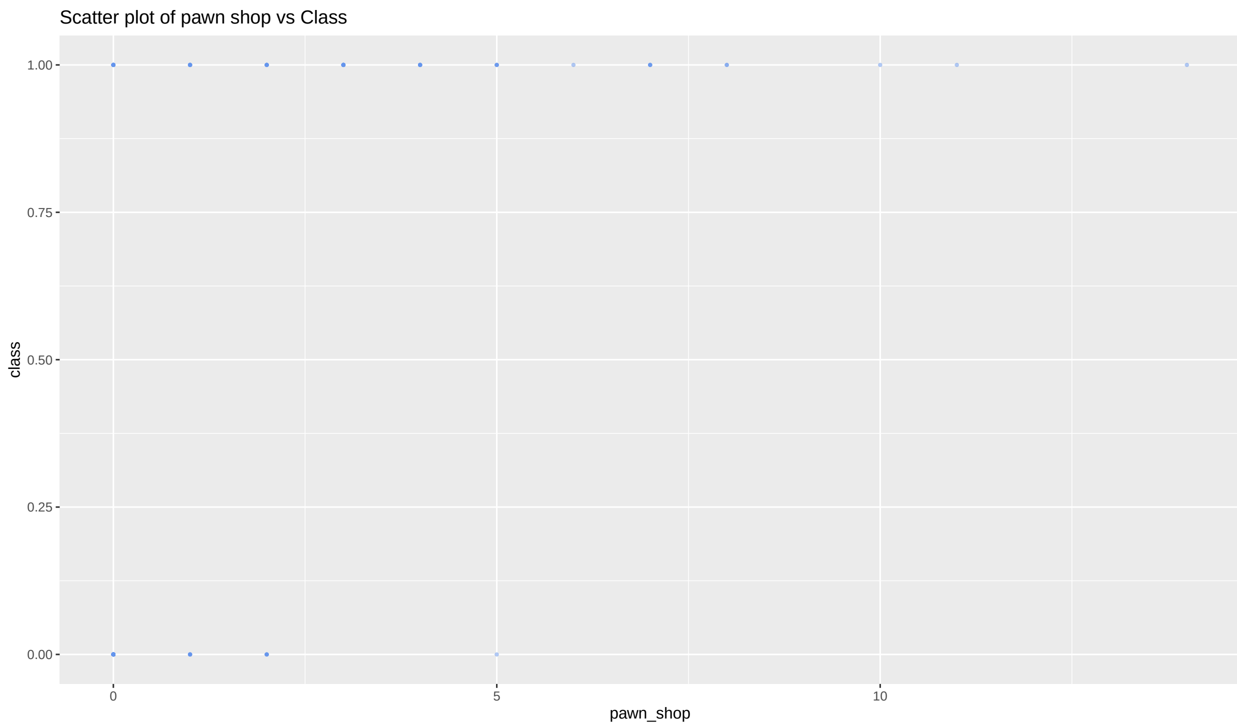


Figure 5: Pawn shop vs dispensaries

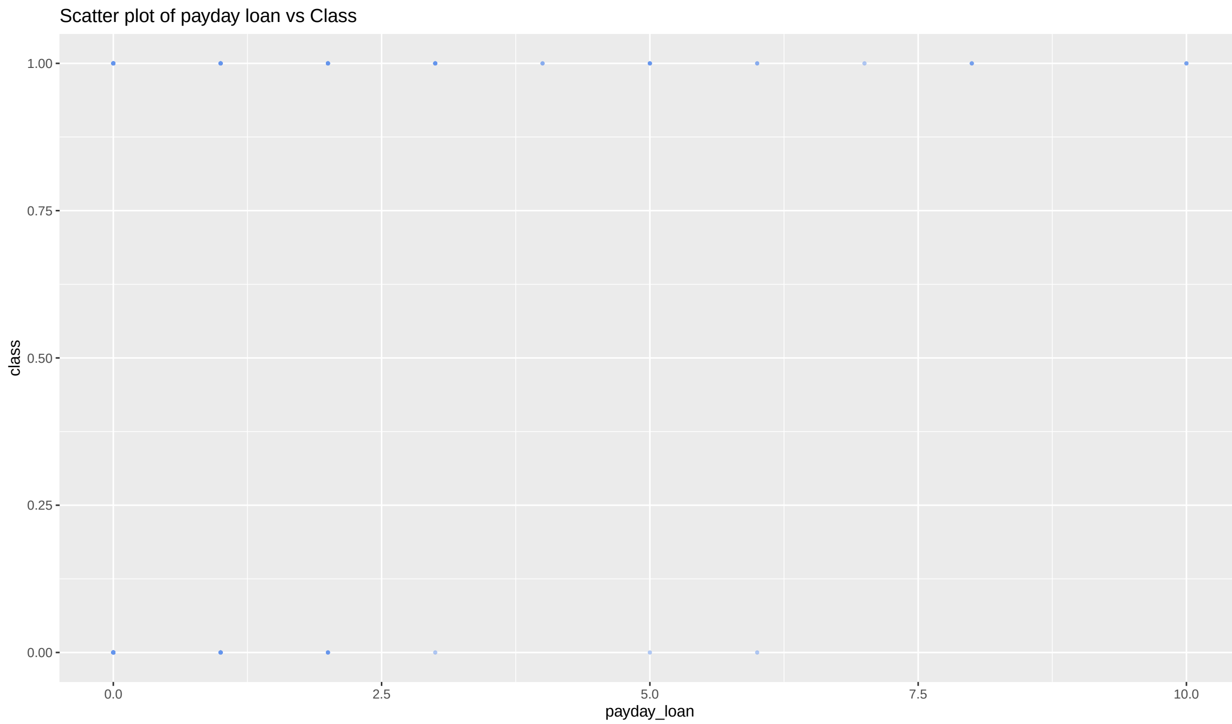


Figure 6: Payday loan vs dispensaries

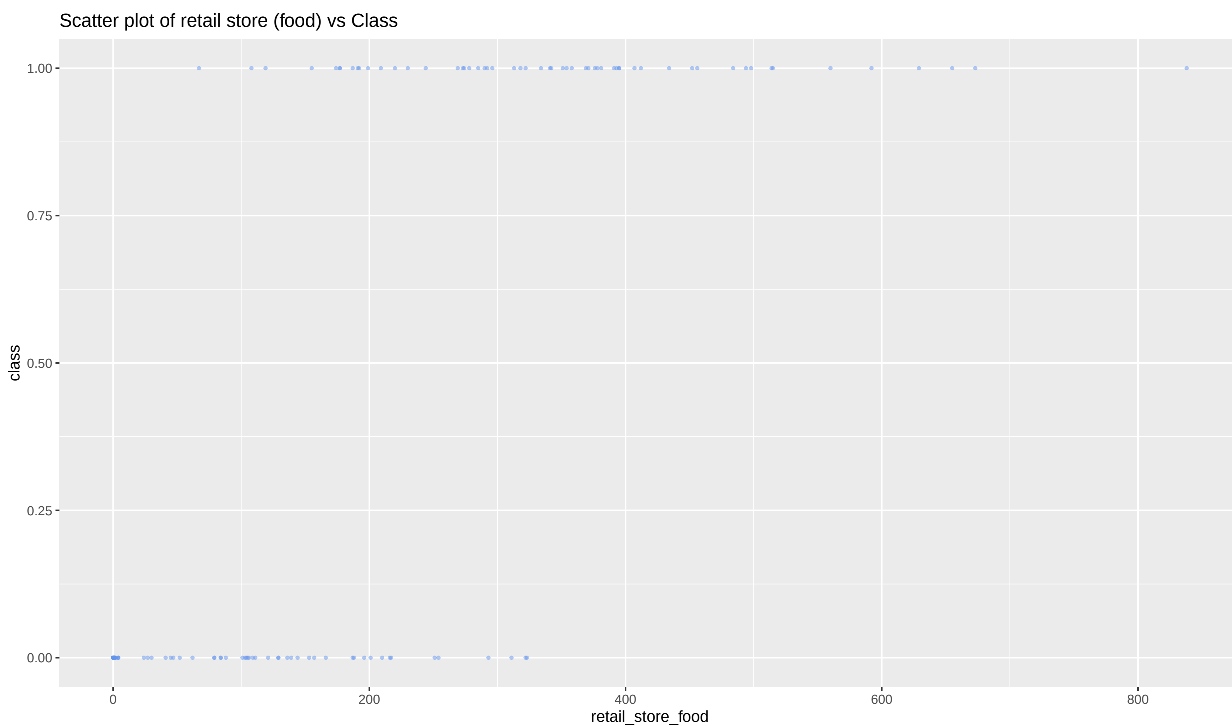


Figure 7: Retail store (food) vs dispensaries

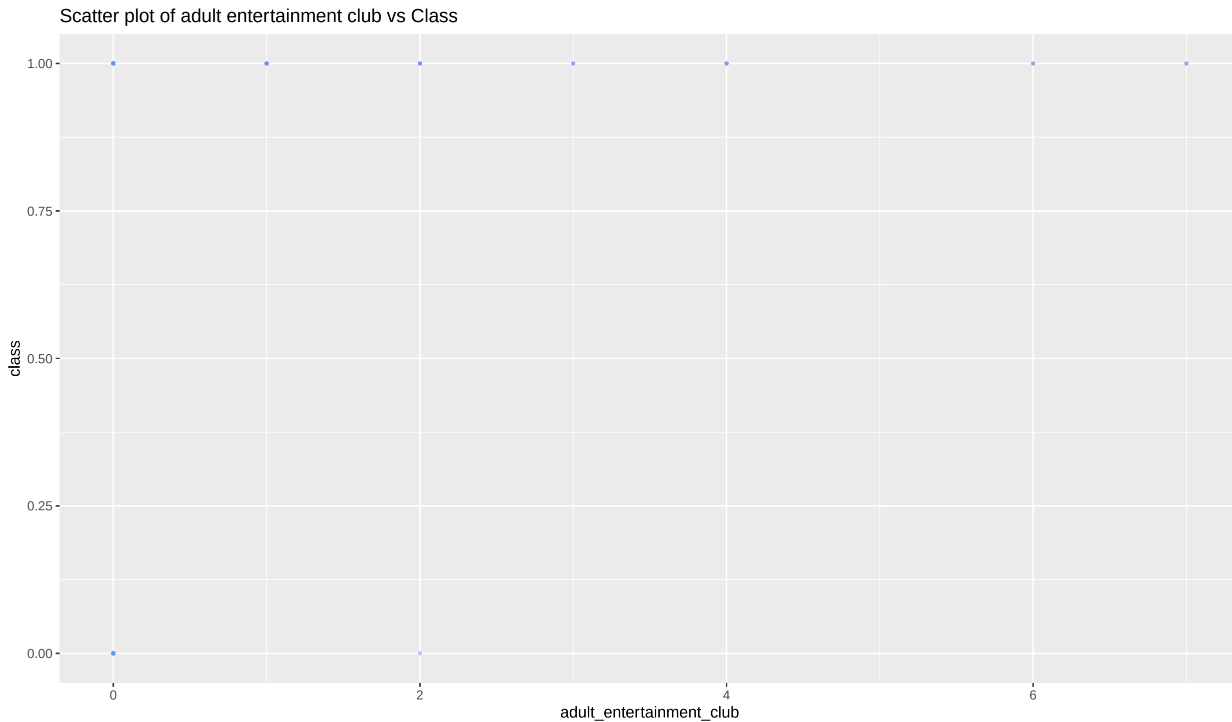


Figure 8: Adult entertainment club vs dispensaries

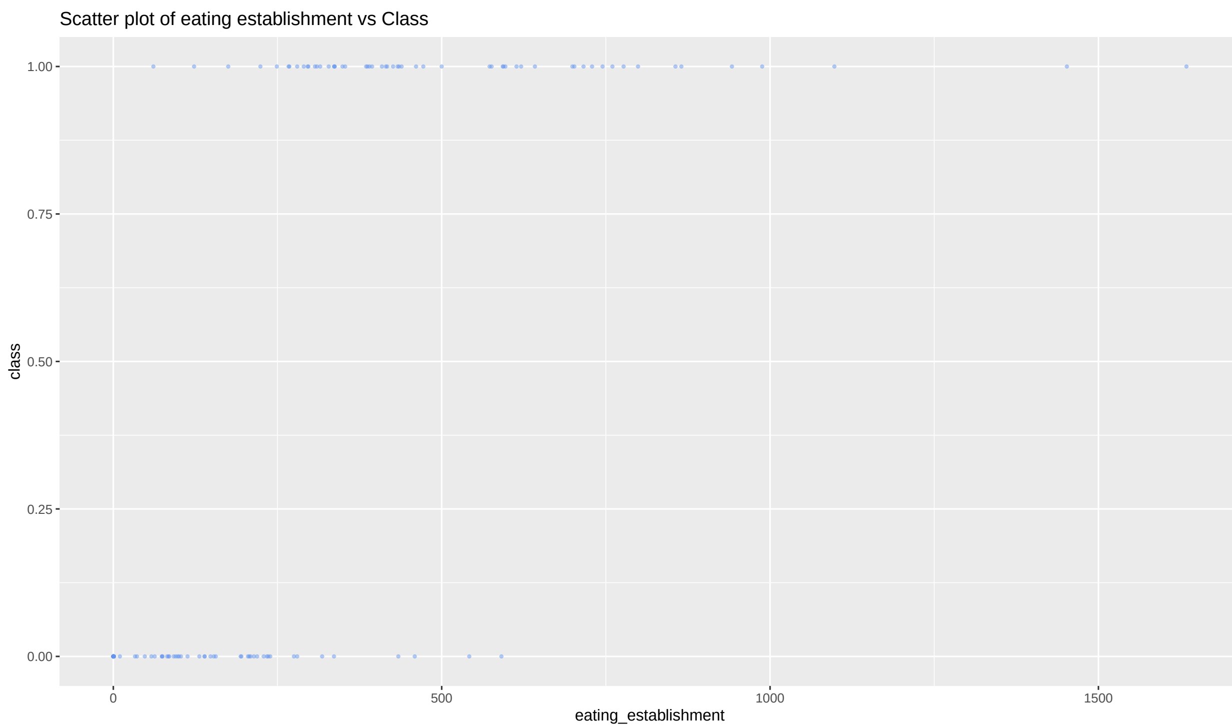
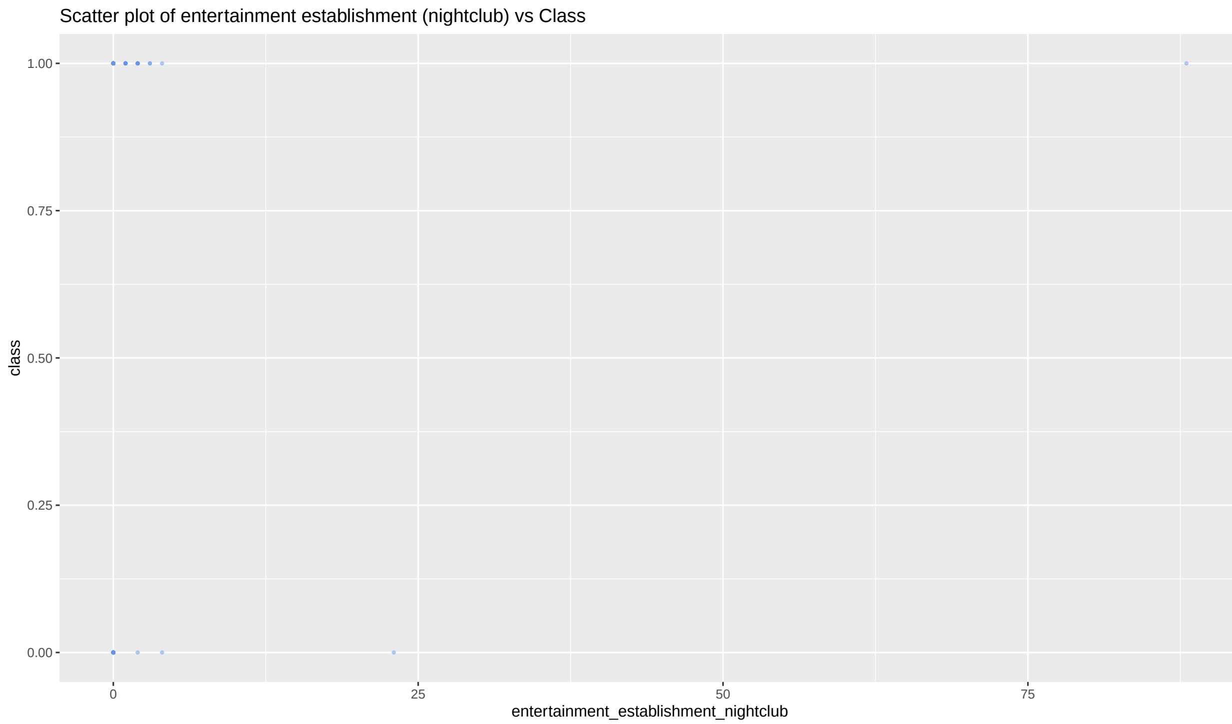


Figure 9: Eating establishment vs dispensaries



Spatial Visualization

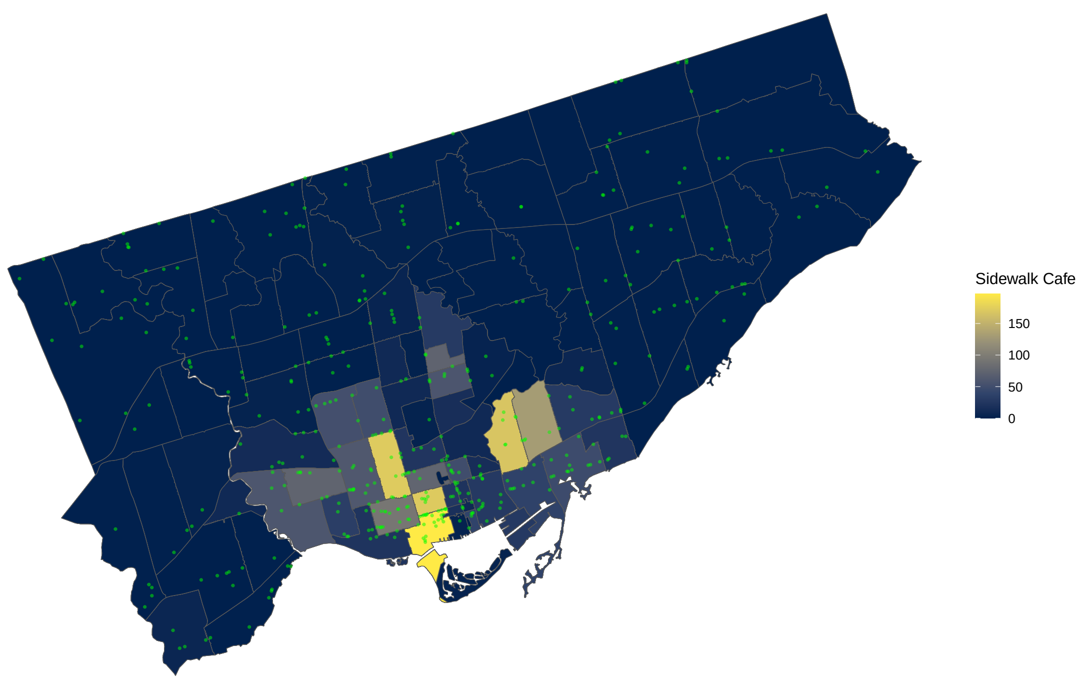


Figure 10: Sidewalk cafe vs dispensaries

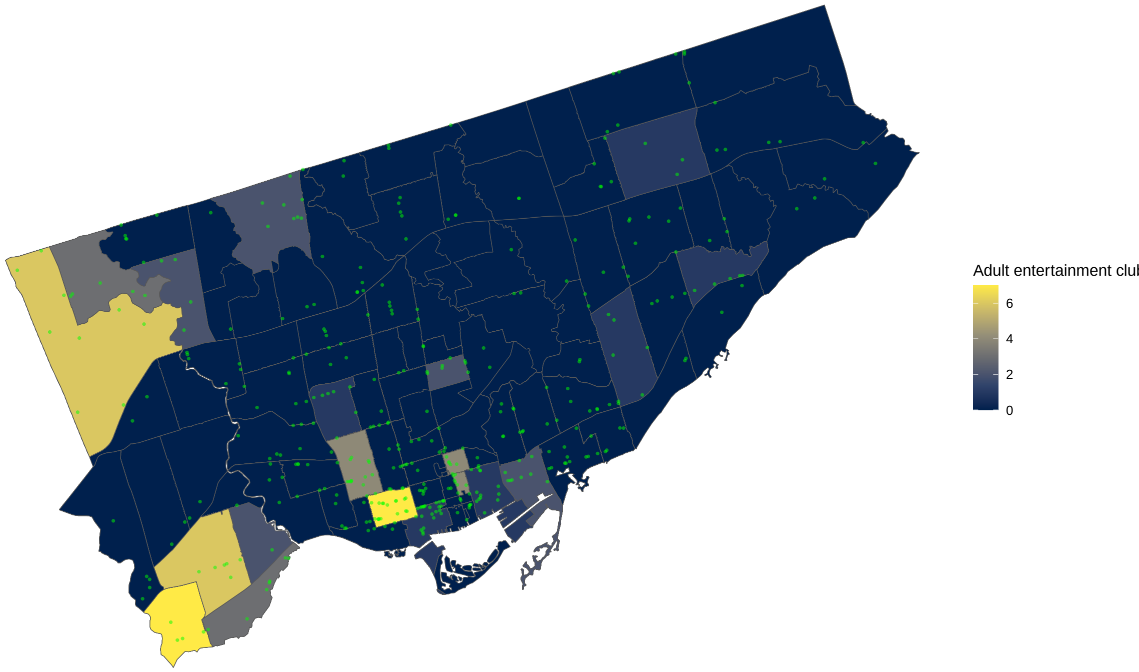
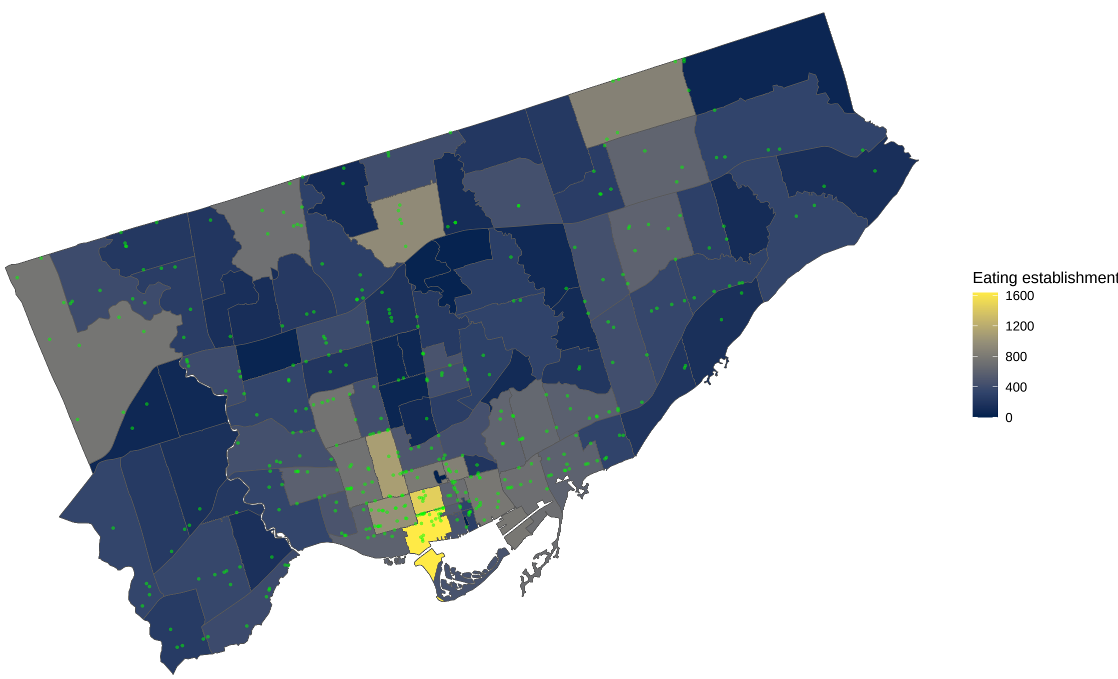
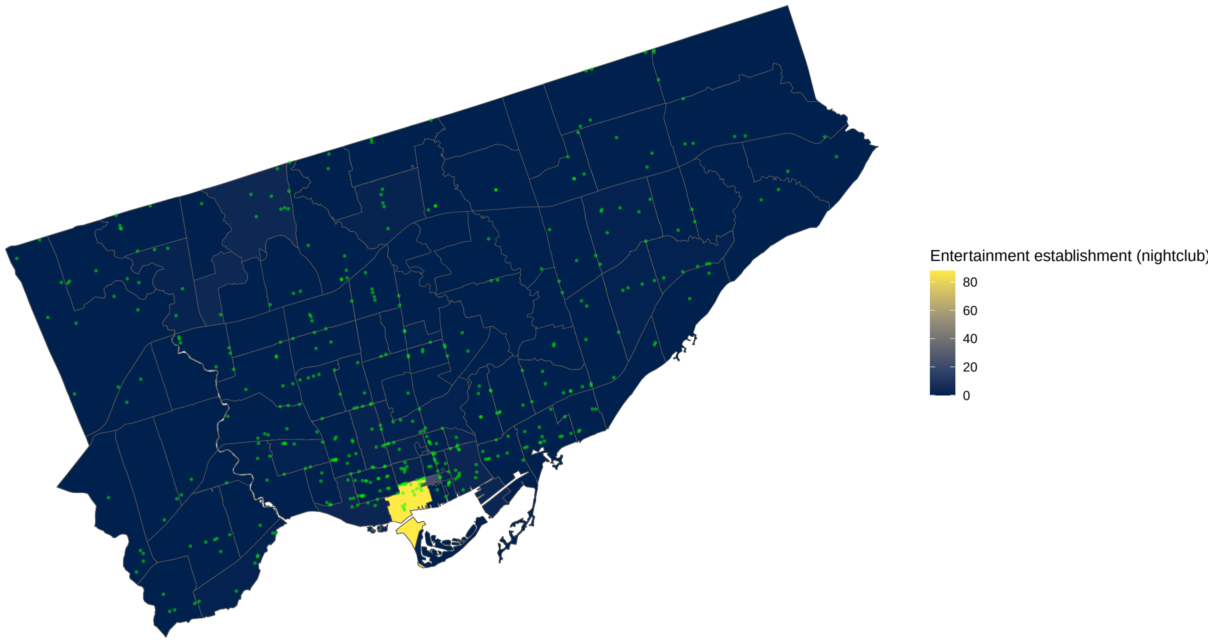


Figure 11: Adult Entertainment Club vs dispensaries





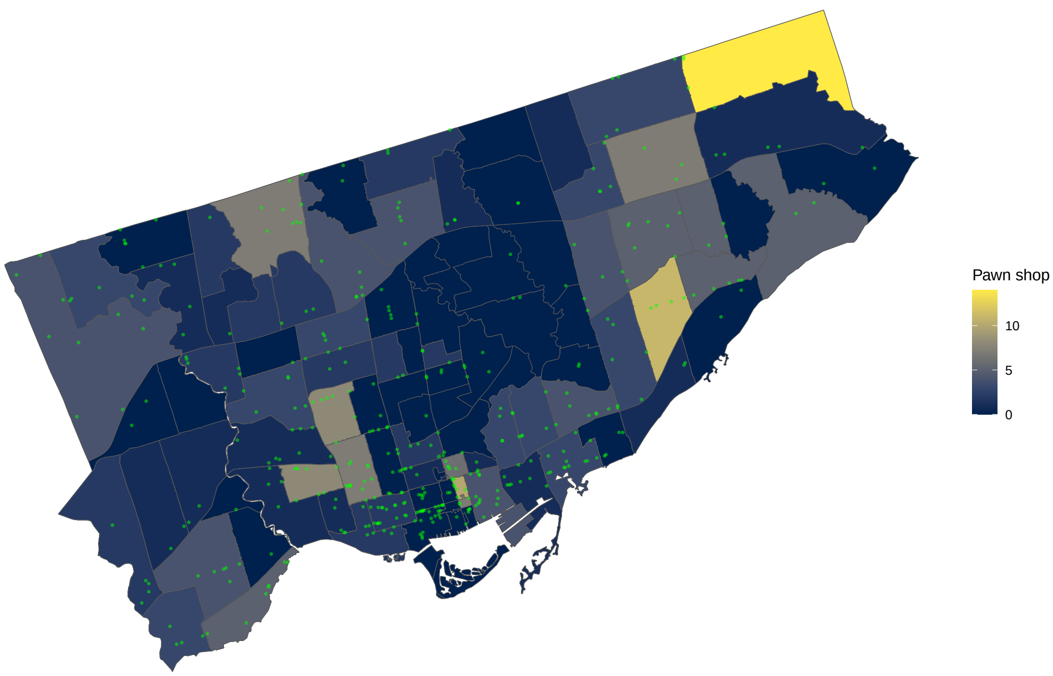


Figure 12: Pawn shop vs dispensaries

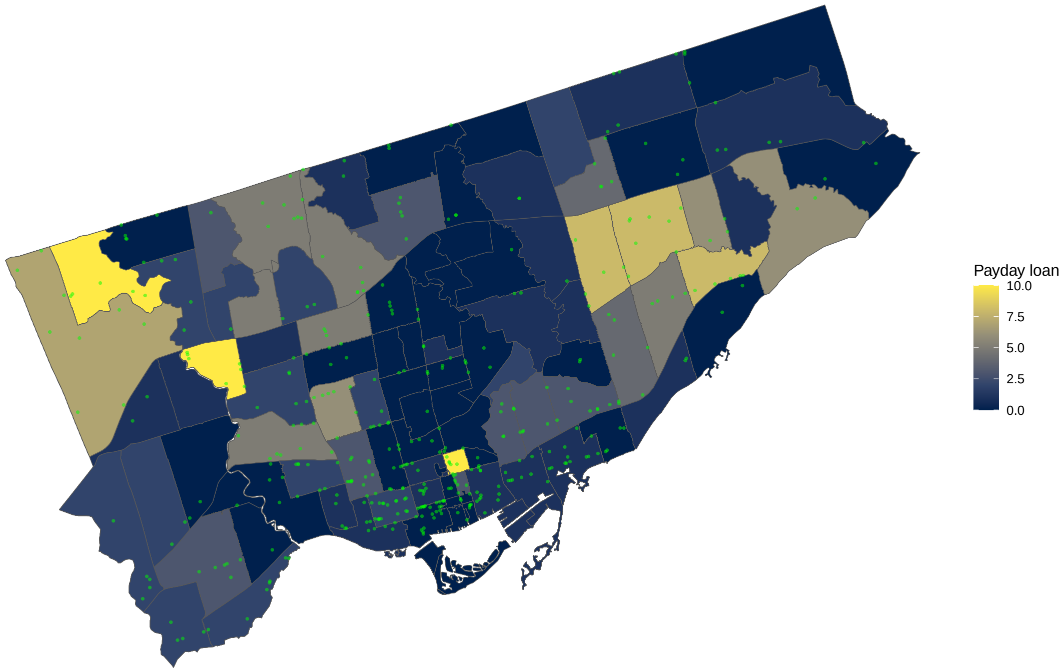


Figure 13: Payday loan vs dispensaries

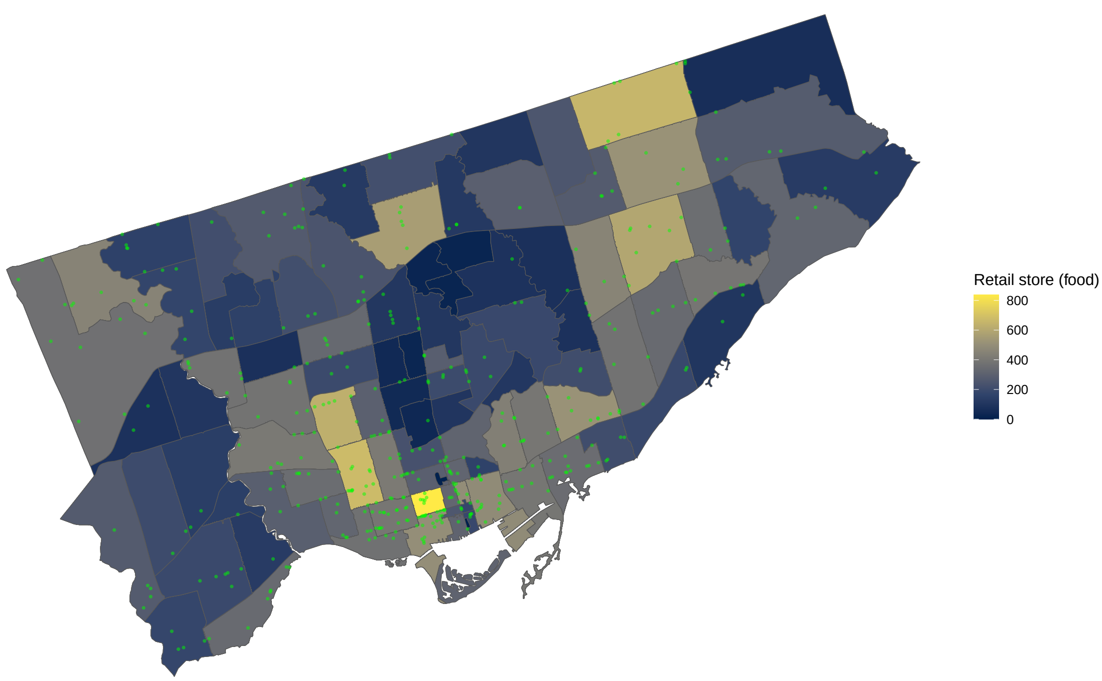


Figure 14: Retail store (food) vs dispensaries

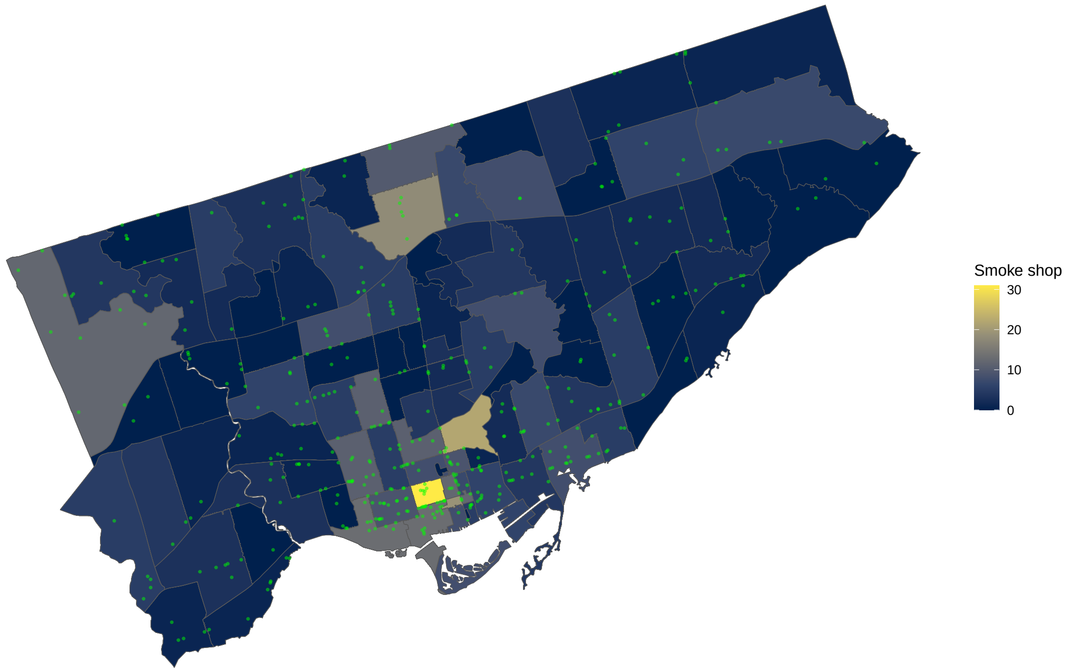


Figure 15: Smoke shop vs dispensaries

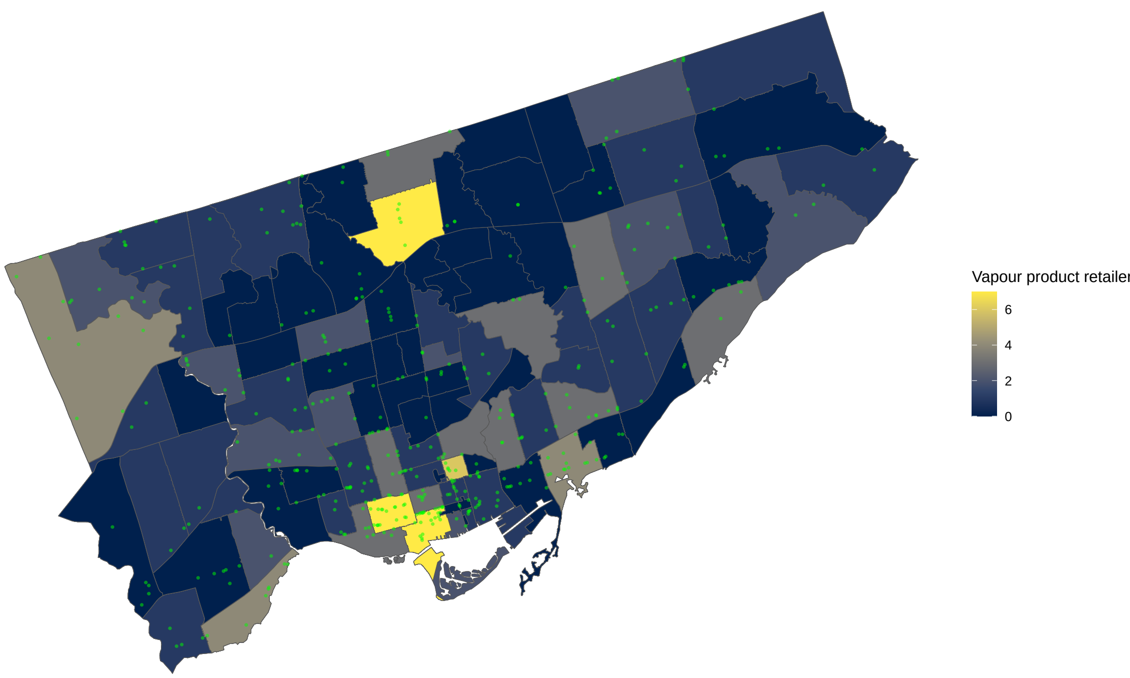


Figure 16: Vapour product retailer vs dispensaries

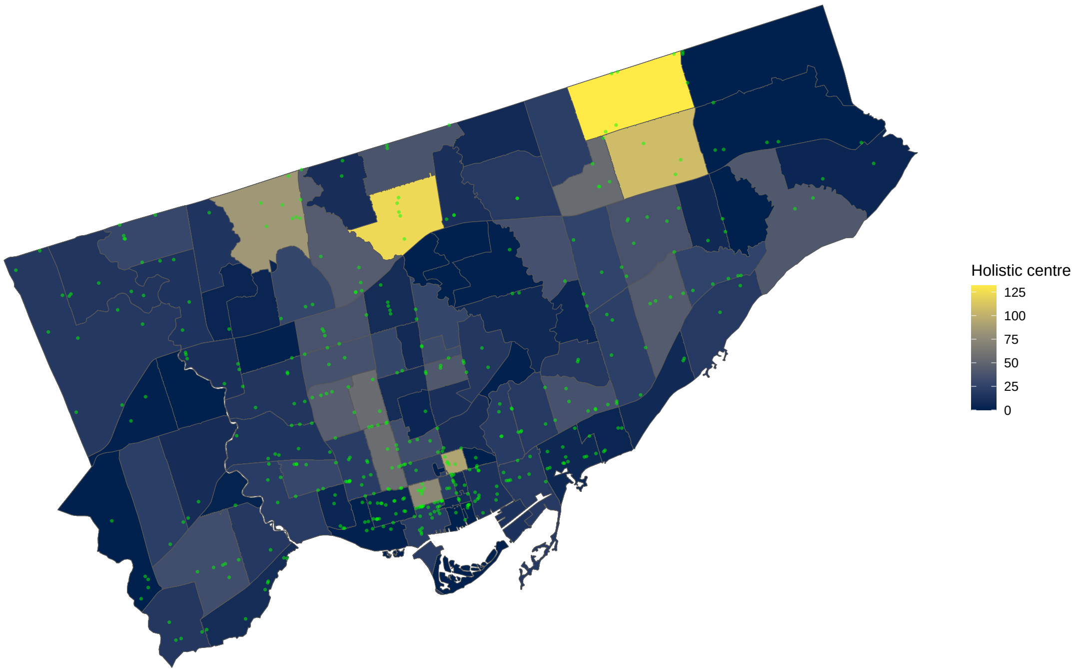


Figure 17: Holistic centre vs dispensaries

Data Summary

Table 1: Data Summary

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
count_dispensaries	15	0	3.6	4.0	0.0	2.0	25.0	
class	2	0	0.5	0.5	0.0	0.0	1.0	
holistic_centre	45	0	18.6	24.7	0.0	11.0	132.0	
eating_establishment	97	0	326.2	303.9	0.0	271.5	1634.0	
entertainment_establishment_nightclub	7	0	1.3	8.4	0.0	0.0	88.0	
pawn_shop	12	0	1.8	2.6	0.0	0.5	14.0	
payday_loan	10	0	1.6	2.4	0.0	1.0	10.0	
sidewalk_cafe	37	0	17.2	37.8	0.0	0.0	197.0	
retail_store_food	96	0	223.7	177.7	0.0	194.0	838.0	
smoke_shop	18	0	3.8	5.1	0.0	2.0	31.0	
vapour_product_retailer	7	0	1.0	1.5	0.0	0.0	7.0	
adult_entertainment_club	7	0	0.5	1.4	0.0	0.0	7.0	

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 = merged_data_wclass1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4452  -0.2712  -0.1414   0.1228   2.5797
##
## Coefficients:
##                                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)                           -4.599240   1.000924  -4.595 4.33e-06 ***
## adult_entertainment_club               1.584667   0.833307   1.902   0.0572 .
## eating_establishment                  -0.003149   0.008553  -0.368   0.7127
## entertainment_establishment_nightclub -0.182940   0.093177  -1.963   0.0496 *
## pawn_shop                              0.561580   0.334643   1.678   0.0933 .
## payday_loan                            0.077889   0.266153   0.293   0.7698
## sidewalk_cafe                          0.094693   0.045760   2.069   0.0385 *
## retail_store_food                      0.012844   0.010616   1.210   0.2263
## smoke_shop                             0.234625   0.146066   1.606   0.1082
## vapour_product_retailer               -0.186916   0.447312  -0.418   0.6760
## holistic_centre                        0.007242   0.028085   0.258   0.7965
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 160.78  on 115  degrees of freedom
## Residual deviance:  50.10  on 105  degrees of freedom
## AIC: 72.1
##
## Number of Fisher Scoring iterations: 8
```

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 = merged_data_wclass1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4183  -0.2753  -0.1479   0.1396   2.4565
##
## Coefficients:
##                                        Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)                         -4.510607   0.946513  -4.766 1.88e-06 ***
## adult_entertainment_club             1.391189   0.763506   1.822   0.0684 .
## entertainment_establishment_nightclub -0.186771   0.085991  -2.172   0.0299 *
## pawn_shop                            0.644741   0.332713   1.938   0.0526 .
## sidewalk_cafe                        0.077921   0.033907   2.298   0.0216 *
## retail_store_food                    0.009958   0.004474   2.226   0.0260 *
## smoke_shop                           0.178091   0.104857   1.698   0.0894 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 160.776  on 115  degrees of freedom
## Residual deviance:  50.621  on 109  degrees of freedom
## AIC: 64.621
##
## Number of Fisher Scoring iterations: 8
```

Exponentiate the coefficients

```
##                           (Intercept)              adult_entertainment_club
##                            0.01099179                            4.01962819
## entertainment_establishment_nightclub                             pawn_shop
##                            0.82963371                            1.90549376
##                         sidewalk_cafe                     retail_store_food
##                            1.08103713                            1.01000763
##                            smoke_shop
##                            1.19493421
```

K-fold cross-validation

```
## Generalized Linear Model
##
## 116 samples
##   6 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 104, 105, 104, 104, 104, 105, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.8871212  0.7720588
```

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 black- and 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 businesses. It would be of particular interest to investigate these businesses 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[1] 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.

[1] As of 2005.