

ACMT Example: Calculating a Neighborhood Deprivation Index

Amy Youngbloom

1/13/2022

Introduction

The ACMT can also be used to generate interpolated measures based on standard neighborhood indices that use American Community Survey data. These indices provide useful and consistent ways for researchers to operationalize measures such as deprivation, social vulnerability, or social fragmentation. One example of an index measure such as this is the Neighborhood Deprivation Index (NDI) developed by Messer et al. (2006). The NDI is a composite measure using socioeconomic variables to describe the level of deprivation of a given area. The NDI relies on American Community Survey variables, all of which are built into the default list of variables in the ACMT. As such the NDI may be a useful tool in comparing interpolated neighborhood deprivation levels around given points of interest.

In this example, we will utilize a list of establishments in Seattle either with an active marijuana retailer license or with an active alcohol license and use the NDI to compare levels of neighborhood depravity around each type of establishment.

We take the following steps:

1. Download dispensary and alcohol retailer data and reformat into lists from WA Liquor and Cannabis Public Record website: <https://lcb.wa.gov/records/frequently-requested-lists>
2. Geocode dispensary and alcohol retailers
3. Create a function to pull ACMT measures
4. Designate the ACS measures relevant to the NDI measure
5. Pull designated ACS measures
6. Transform list of measures into a wide dataframe
7. Run a principal component analysis to find weights for each variable measure
8. Calculate the NDI measure
9. Create a boxplot to compare the NDI measures for dispensaries versus alcohol retailers
10. Calculate a t-test to statistically compare NDI measures

```
source("setup-acmt.R")
library(readxl)
library(tidyverse)
library(janitor)
```

1. Download data and reformat into lists

Verify the urls below for the Dispensaries and Alcohol retailers below. The lists are updated regularly and can be found here: <https://lcb.wa.gov/records/frequently-requested-lists>. See urls for these lists: Cannabis License Applicants and Off Premises, Licensees

```

download.file(url = "https://lcb.wa.gov/sites/default/files/publications/Public_Records/2019/CannabisAppl
destfile = "external_data/downloaded_cannabis.xls") ##check records website if url is out

download.file(url = "https://lcb.wa.gov/sites/default/files/publications/Public_Records/2019/Off%20Prem
destfile = "external_data/downloaded_alcohol.xlsx")

cannabis_dataframe <- read_excel("external_data/downloaded_cannabis.xls")
alcohol_dataframe<-read_excel("external_data/downloaded_alcohol.xlsx")

#limit to Seattle only dispensaries and create address field:
seattle_dispensary<-cannabis_dataframe %>%
  janitor::clean_names() %>%
  filter(city == "SEATTLE", priv_desc == "MARIJUANA RETAILER", (privilege_status=="ACTIVE (ISSUED)" | p
  mutate(full_address = paste(street_address,
                              city, state,
                              zip_code, sep = ", "), ID=paste(as.character(row_number()), 'disp', sep='.'),
          status=privilege_status,
          license_type='dispensary') %>%
  subset(select=c(ID, tradename, status, full_address, license_type))

seattle_alcohol<-alcohol_dataframe %>%
  janitor::clean_names() %>%
  filter(loc_city == "SEATTLE", status == "ACTIVE (ISSUED)" | status == "PENDING (ISSUED)",
         privilege != "BEER/WINE GIFT DELIVERY") %>%
  mutate(full_address = paste(loc_address,
                              loc_city, loc_st,
                              loc_zip, sep = ", "), ID=paste(as.character(row_number()), 'alcohol', sep
          license_type='alcohol_offpremise') %>%
  subset(select=c(ID, tradename, status, full_address, license_type, privilege))

#transform into lists & merge:
seattle_dispensary_list<-as.list(seattle_dispensary$full_address)
seattle_dispensary_list<-structure(seattle_dispensary_list, names=(seattle_dispensary$ID))

seattle_alcohol_list<-as.list(seattle_alcohol$full_address)
seattle_alcohol_list<-structure(seattle_alcohol_list, names=(seattle_alcohol$ID))

seattle_alcohol_dispensary_list<-c(seattle_dispensary_list[1:20], seattle_alcohol_list[1:20])

head(seattle_alcohol_dispensary_list)

## $`1.disp`
## [1] "224 NICKERSON ST, SEATTLE, WA, 981091622"
##
## $`2.disp`
## [1] "5209 MARTIN LUTHER KING JR., SEATTLE, WA, 981186131"
##
## $`3.disp`
## [1] "321 NE 45TH ST, SEATTLE, WA, 981056133"
##
## $`4.disp`
## [1] "5300 17TH AVE NW STE A, SEATTLE, WA, 981073815"
##

```

```
## $`5.disp`
## [1] "2214 1ST AVE, SEATTLE, WA, 981211615"
##
## $`6.disp`
## [1] "5440 CALIFORNIA AVE SW, SEATTLE, WA, 981361513"
```

2. Geocode dispensary and alcohol retailers list

```
seattle_lat_to_long_disp_alcohol_list<-lapply(seattle_alcohol_dispensary_list, geocode)
print(seattle_lat_to_long_disp_alcohol_list[1])
```

```
## $`1.disp`
## $`1.disp`$latitude
## [1] 47.64762
##
## $`1.disp`$longitude
## [1] -122.3528
```

3. Create a function to pull ACMT measures

Once we have a geocoded list of dispensaries and off premise alcohol sale licensees, we can write a function to pull ACS measures that we will need to calculate the NDI.

```
get_variable_measures_from_acmt <- function(dispensary_to_lat_long_list, radius_vector, year, names_of_variable_to_get) {
  dispensary_to_radius_to_variable_to_measures_list <- vector(mode="list", length=length(dispensary_to_lat_long_list))
  names(dispensary_to_radius_to_variable_to_measures_list) <- names(dispensary_to_lat_long_list)

  for(dispensary in names(dispensary_to_radius_to_variable_to_measures_list)) {
    radius_to_variable_to_measures_list <- vector(mode="list", length=length(radius_vector))
    names(radius_to_variable_to_measures_list) <- as.character(radius_vector)
    for (radius in radius_vector) {
      print(dispensary)
      print(radius)

      latitude <- dispensary_to_lat_long_list[[dispensary]]$latitude
      longitude <- dispensary_to_lat_long_list[[dispensary]]$longitude

      environmental_measures <- get_acmt_standard_array(long=longitude,
                                                         lat=latitude, radius_meters = radius, year=year)

      variable_to_measures_list <- vector(mode="list", length=length(names_of_variable_to_get))
      names(variable_to_measures_list) <- names_of_variable_to_get
      for (name_of_variable in names_of_variable_to_get) {
        value_of_variable <- environmental_measures[environmental_measures$names == name_of_variable, ]
        variable_to_measures_list[[name_of_variable]] <- value_of_variable
      }
      radius_to_variable_to_measures_list[[which(radius == radius_vector)]] <- variable_to_measures_list
      radius_to_variable_to_measures_list[[as.character(radius)]] <- variable_to_measures_list
    }
  }
  dispensary_to_radius_to_variable_to_measures_list[[dispensary]] <- radius_to_variable_to_measures_list
}
```



```

      radius_vector <- c(radius_vector, as.numeric(radius_character))
      variable_vector <- c(variable_vector, variable)
      value_vector <- c(value_vector, seattle_dispensary_ndi_measures[[dispensary]][[radius_character]])
    }
  }
}

dataframe_dispensary <- data.frame(dispensary_vector, radius_vector, variable_vector, value_vector,
names(dataframe_dispensary) <- c("dispensary", "radius", "variable", "value")
return(dataframe_dispensary)
}

dataframe_dispensary <- convert_to_dataframe(seattle_dispensary_ndi_measures)

dataframe_dispensary$radius<-NULL
dispensary_wide_dataframe<-reshape(dataframe_dispensary, v.names="value",timevar = "variable", idvar = "dispensary")

seattle_dispensary$privilege<-'marijuana retailer'
seattle_alcohol_dispensary<-rbind(seattle_dispensary, seattle_alcohol)

seattle_disp_alcohol_ndi_measures_wide<-merge(dispensary_wide_dataframe, seattle_alcohol_dispensary, by = "dispensary")

```

7. Run a principal component analysis to find weights for each variable measure

Now we can use the ACS measures that were just pulled to calculate the NDI score for the neighborhood around each dispensary. First we need to combine measures to calculate the total percent of residents with professional degrees and the total percent of housing that is crowded (2.01 or greater occupants per room), and total percent of households with an income less than \$30,000.

To create the composite NDI value, we first must get the weight for individual measure by conducting a principal component analysis of the eight designated NDI variables: (1) percent of males in professional occupations, (2) percent of households in poverty, (3) percent of residents with no high school diploma, (4) the percent of residents who are unemployed, (5) the percent of crowded housing, (6) the percent of households with income less than \$30,000, (7) the percent of female-headed households with dependents less than 18 years old, and (8) the percent of households on public assistance. The loadings generated by the PCA will be used to weight each measure in calculating the overall NDI summary measure for each tract. The resulting composite NDI measure is multiplied by -1 so that a higher NDI measure indicates a neighborhood with a higher level of deprivation. Finally, we standardize the NDI measures by subtracting the mean and dividing by the standard deviation so that the measure has a mean of 0 and a standard deviation of 1.

```

#clean-up NDI measures
seattle_dispensary_ndi<-seattle_disp_alcohol_ndi_measures_wide %>%
  mutate(total_males_management_professional = value.males_in_management_count+value.males_in_professional_count,
         total_crowded_housing = value.owner_2.01_or_more_per_room_count
         + value.renter_2.01_or_more_per_room_count,
         total_income_below_30k = value.household_income_25_29k_count + value.household_income_20_24k_count)
  mutate(percent_males_management_professional = total_males_management_professional / value.males_16_or_over_count,
         percent_crowded_housing = total_crowded_housing / (value.total_owner_occ_housing_units_rooms_count+value.total_renter_occ_housing_units_rooms_count),
         percent_income_below_30k = total_income_below_30k/value.households_income_determined_count)

#PCA Factor Analysis
NDI_measures<-seattle_dispensary_ndi %>%
  subset(select=c(percent_males_management_professional, value.households_in_poverty_proportion, value.households_in_public_assistance, value.percent_unemployed, value.percent_female_headed_households_with_dependents_less_than_18_years_old, value.percent_households_on_public_assistance, value.percent_households_with_income_less_than_30k, value.percent_households_in_poverty))

```

```
ndi_pca<-princomp(NDI_measures, cor=TRUE)
```

8. Calculate the NDI measure

Once the NDI score has been calculated and standardized, we can construct a boxplot to compare the mean and distribution of NDI scores for establishments with a marijuana retailer license to those with an off-premise alcohol license.

```
##assign loading values for each variable
percent_males_management_professional_loading<-ndi_pca$loadings[1]
value.households_in_poverty_proportion_loading<-ndi_pca$loadings[2]
value.no_hsdiploma_proportion_loading<-ndi_pca$loadings[3]
value.unemployed_proportion_loading<-ndi_pca$loadings[4]
percent_crowded_housing_loading<-ndi_pca$loadings[5]
percent_income_below_30k_loading<-ndi_pca$loadings[6]
value.female_head_kids_proportion_loading<-ndi_pca$loadings[7]
value.public_asst_proportion_loading<-ndi_pca$loadings[8]

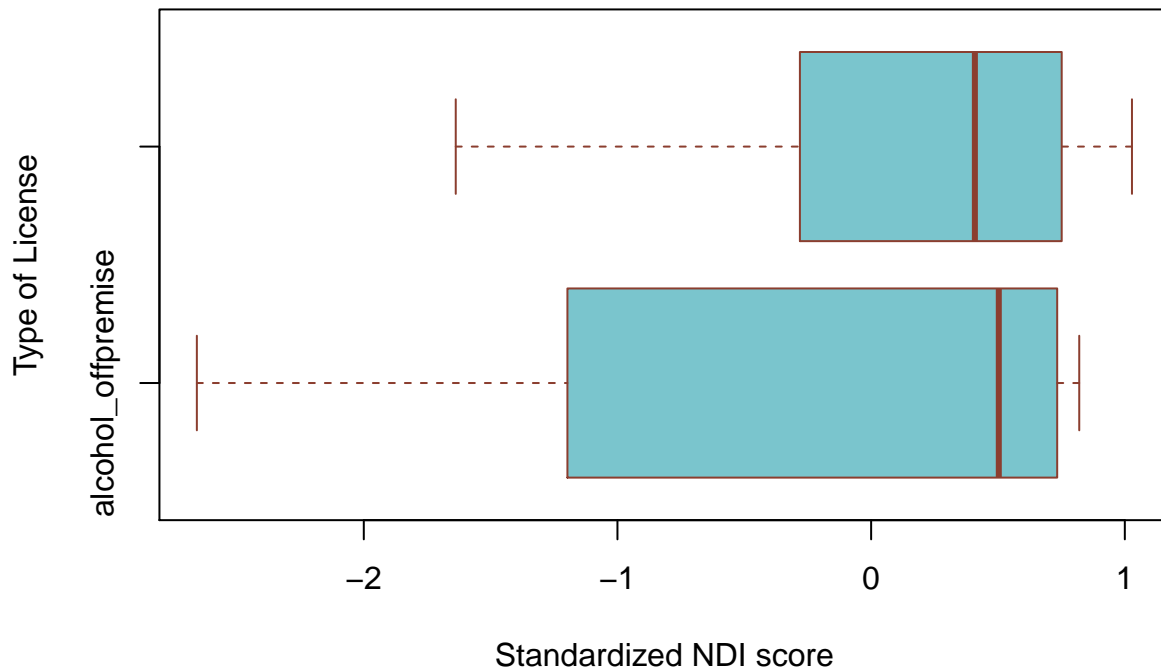
#Calculated & standardize weighted NDI value using pca loadings
seattle_dispensary_ndi <-seattle_dispensary_ndi%>%
  mutate(ndi_value=((percent_males_management_professional*percent_males_management_professional_loading+
    (value.households_in_poverty_proportion*value.households_in_poverty_proportion_loading)+
    (value.no_hsdiploma_proportion*value.no_hsdiploma_proportion_loading)+
    (value.unemployed_proportion*value.unemployed_proportion_loading)+
    (percent_crowded_housing*percent_crowded_housing_loading)) +
    (percent_income_below_30k*percent_income_below_30k_loading) +
    (value.female_head_kids_proportion*value.female_head_kids_proportion_loading)+
    (value.public_asst_proportion*value.public_asst_proportion_loading)*-1) %>%
  mutate(ndi_standardized=(ndi_value-mean(ndi_value))/sd(ndi_value)) %>%
  mutate(dispensary_active=ifelse(status=="ACTIVE (ISSUED)" | status=="PENDING (ISSUED)", 1, 0))
```

9. Create a boxplot to compare the NDI measures for dispensaries versus alcohol retailers

Once the NDI score has been calculated and standardized, we can construct a boxplot to compare the mean and distribution of NDI scores for establishments with a marijuana retailer license to those with an off-premise alcohol license.

```
#look at specific types of alcohol privileges
boxplot(ndi_standardized~license_type,
  data=seattle_dispensary_ndi,
  main = "Neighborhood Deprivation Scores",
  xlab = "Standardized NDI score",
  ylab = "Type of License",
  col = "cadetblue3",
  border= "coral4",
  horizontal=TRUE)
```

Neighborhood Deprivation Scores



10. Calculate a t-test to statistically compare NDI measures

We can also statistically compare the NDI measures of neighborhoods around alcohol and marijuana retailers using an independent t-test.

```
ndi_ttest<-t.test(seattle_dispensary_ndi$ndi_value~seattle_dispensary_ndi$license_type, var.equal=TRUE)
ndi_ttest
```

```
##
## Two Sample t-test
##
## data: seattle_dispensary_ndi$ndi_value by seattle_dispensary_ndi$license_type
## t = -0.78399, df = 38, p-value = 0.4379
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.05815909 0.02568758
## sample estimates:
## mean in group alcohol_offpremise      mean in group dispensary
## -0.004679553                        0.011556198
```

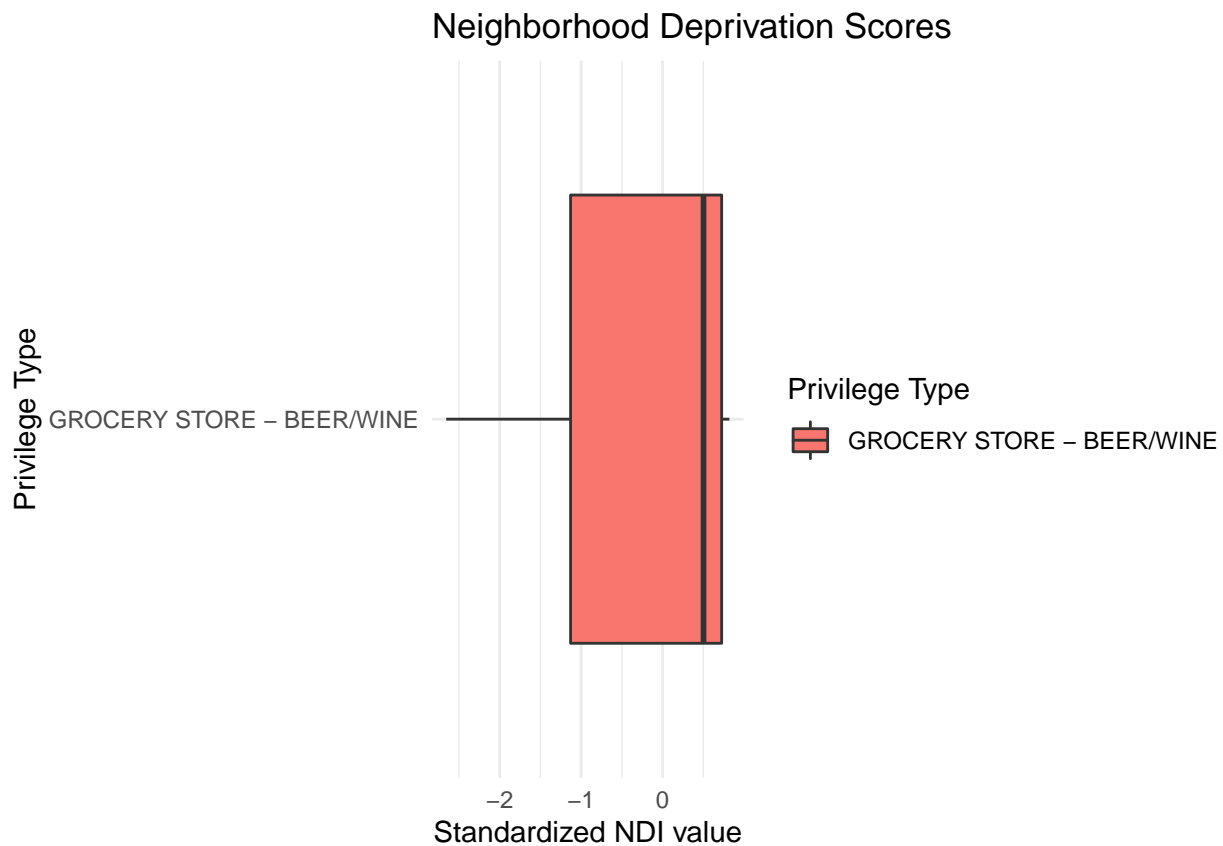
Additional boxplot

In addition to the comparisons of marijuana and alcohol retailer NDI measures, we can also compare across multiple types of alcohol retailers by first constructing a boxplot by license type for alcohol retailers.

```
ndi_privilege_boxplot<- ggplot(seattle_dispensary_ndi[seattle_dispensary_ndi$license_type=='alcohol_offpremise',])
  geom_boxplot()+
  coord_flip()+
```

```
labs(title="Neighborhood Deprivation Scores", y="Standardized NDI value", x="Privilege Type", fill="P")
theme_minimal()
```

```
ndi_privilege_boxplot
```



Additional statistical analysis

We can also look at statistical differences between the deprivation level of neighborhoods around each alcohol establishment using an ANOVA test. First we will look at the mean and sd for NDI scores by privilege type.

```
#look at table of means:
ndi_mean_table<-seattle_dispensary_ndi[seattle_dispensary_ndi$license_type=='alcohol_offpremise',] %>%
  group_by(privilege) %>%
  summarise_at(vars(ndi_standardized), list(NDI_mean=mean, NDI_SD=sd))

ndi_aov<-aov(seattle_dispensary_ndi$ndi_standardized~seattle_dispensary_ndi$privilege, data=seattle_disp)
print(ndi_mean_table)
```

```
## # A tibble: 1 x 3
##   privilege          NDI_mean NDI_SD
##   <chr>             <dbl>   <dbl>
## 1 GROCERY STORE - BEER/WINE -0.125  1.15
```

```
summary(ndi_aov)
```

```
##
##               Df Sum Sq Mean Sq F value Pr(>F)
## seattle_dispensary_ndi$privilege  1    0.62  0.6208    0.615  0.438
```


Residuals 38 38.38 1.0100

References

Messer, L.C., Laraia, B.A., Kaufman, J.K., Eyster, J., Holzman, C., Culhane, J., Elo, I., Burke, J.G., & O'Campo, P. (2006). The Development of a Standardized Neighborhood Deprivation Index. *Journal of Urban Health*, 84(6): 1041-1062.