ACMT Example: Population density over distance (advanced)

Weipeng Zhou

2/22/2021

Introduction

Automatic Context Measurement Tool (ACMT) is a convenient tool for studying neighbourhoods in the United States. Based on the user-provided address and radius, ACMT locates a geographical area and outputs context measurements (population, education level, commute time, etc.) for the area. ACMT is easy to install, highly reproducible and works consistently across computer platforms.

There are various ways of using ACMT and here we show one example – 5-city comparison of population density decay as function of distance from City Hall.

We use ACMT to get population densities for 5 cities (Seattle, Los Angeles, Chicago, New York and Boston) over 5 radiuses (1000, 2000, 3000, 4000, 5000). We consider the center of a city to be the location of its City Hall. We compute the density using the land area within a region only. After getting the densities, we make a plot of density over radius, categoried by cities; and we will be able to tell which city is the most populated across space. We also explore the relationship between population density and national walkability index.

We take the following steps:

1. Find the addresses of each city's City Hall
2. Use ACMT's geocoder to convert the addresses to coordinates
3. Use ACMT to get the population, land area and national walkability index measurement for each coordinate over 5 radiuses (1000, 2000, 3000, 4000, 5000)
4. Calculate population density
5. Plot density vs. radius, categoried by cities, as well as vs. walkability index.

Example

1. Find the addresses of each city's City Hall

We have gathered the City Hall addresses for Seattle, Los Angeles, Chicago, New York and Boston from Google.

```
source("~/workspace/setup-acmt.R")

library(ggplot2)
city_hall_to_address_list <- list(
  seattle_city_hall="600 4th Ave, Seattle, WA 98104",
  los_angeles_city_hall="200 N Spring St, Los Angeles, CA 90012",
  chicago_city_hall="121 N LaSalle St, Chicago, IL 60602",
  new_york_city_hall="City Hall Park, New York, NY 10007",
  boston_city_hall="1 City Hall Square #500, Boston, MA 02201"
)
```

2. Use ACMT's geocoder to convert the addresses to coordinates

ACMT comes with a handy geocoder that converts addresses to latitude/longtitude coordinates. We check if geocoder is available in the version of ACMT you installed; if it is not available, we use the pre-computed coordinates.

```
convert_address_to_lat_long <- function (city_hall_to_address_list) { # function to get get lat/long for each add
ress
  city_hall_to_lat_long_list <- vector(mode="list", length=length(city_hall_to_address_list))
  names(city_hall_to_lat_long_list) <- names(city_hall_to_address_list)
  for (name in names(city_hall_to_address_list)){
    city_hall_to_lat_long_list[[name]] <- geocode(city_hall_to_address_list[[name]])
  }
  return(city_hall_to_lat_long_list)
}
```

```
geocoder_is_available<-as.data.frame(city_hall_to_address_list) %>% t() %>% as.data.frame() %>%
  rename(address=V1) %>%
  mutate(city_hall=row.names(.), geocoder_is_available=FALSE)

city_hall_lat_long<-list(seattle_city_hall=list(latitude=47.60328, longitude=-122.3302),
    los_angeles_city_hall=list(latitude=34.05397, longitude=-118.2436),
    chicago_city_hall=list(latitude=41.88334, longitude=-87.63229),
    new_york_city_hall=list(latitude=40.66392, longitude=-73.93835),
    boston_city_hall=list(latitude=42.35773, longitude=-71.05919))

for(address in geocoder_is_available$address){
  tryCatch({geocode(address)
    geocoder_is_available$geocoder_is_available[geocoder_is_available$address==address]=TRUE}, error=function(con
dition){
    print(condition$message)
    print("Geocoder not available: using stored address to lat/long mappings instead")
  })
}

# call geocoder if available, use hard coded info otherwise
city_hall_to_lat_long_list <- NULL
for(address in geocoder_is_available$address){
  city_hall_name<-geocoder_is_available$city_hall[geocoder_is_available$address==address]
  if(geocoder_is_available$geocoder_is_available[geocoder_is_available$address==address]==TRUE){

    city_hall<-list(address)
    names(city_hall)<-city_hall_name
    city_hall_to_lat_long_list<-append(city_hall_to_lat_long_list, convert_address_to_lat_long(city_hall))
                                                                }
  else(city_hall_to_lat_long_list<-append(city_hall_to_lat_long_list, city_hall_lat_long[names(city_hall_lat_lon
g)==city_hall_name]))
}
```

```
print(city_hall_to_lat_long_list[1])
```

```
## $seattle_city_hall
## $seattle_city_hall$latitude
## [1] 47.60328
##
## $seattle_city_hall$longitude
## [1] -122.3302
##
## $seattle_city_hall$rating
## [1] 0
```

3. Use ACMT to get the population measurement for each coordinate over 5 radiuses (1000, 2000, 3000, 4000, 5000)

We create a function for querying ACMT measurements for a list of coordinates and radiuses. We are interested in the variable `total_pop_count`.

```
# function to get the environmental measures for the variables we are intersted
get_variable_measures_from_acmt <- function (city_hall_to_lat_long_list, radius_vector, year, names_of_variable_to_get, external_data_name_to_info_list=NULL, codes_of_acs_variables_to_get=NA) {
  city_hall_to_radius_to_variable_to_measures_list <- vector(mode="list", length=length(city_hall_to_lat_long_list))
  names(city_hall_to_radius_to_variable_to_measures_list) <- names(city_hall_to_lat_long_list)

  for(city_hall in names(city_hall_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(city_hall)
      print(radius)
      # get lat/long
      latitude <- city_hall_to_lat_long_list[[city_hall]]$latitude
      longitude <- city_hall_to_lat_long_list[[city_hall]]$longitude

      # get environmental measures for all variables
      environmental_measures <- get_acmt_standard_array(long=longitude, lat=latitude, radius_meters = radius, year=year, external_data_name_to_info_list=external_data_name_to_info_list, codes_of_acs_variables_to_get=codes_of_acs_variables_to_get)

      # get environmental measures for the variables are interested
      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$names == name_of_variable, ]$values  # 66370.01, seattle, r=2000, year=2017
        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
    }
    city_hall_to_radius_to_variable_to_measures_list[[city_hall]] <- radius_to_variable_to_measures_list
  }
  return(city_hall_to_radius_to_variable_to_measures_list)
}
```

```
setwd('~/workspace')
city_hall_to_lat_long_list <- city_hall_to_lat_long_list
radius_vector <- c(1000, 2000, 3000, 4000, 5000)
year <- 2017
names_of_variable_to_get <- c("total_pop_count", "AC_UNPR", "NatWalkInd")  # population, land area in acres, national walkability index
codes_of_acs_variables_to_get <- c("B01001_001")  # speed up; ask ACMT to only query this variable from the ACS server to speed up computation; B01001_001 is population; check ACSColumns.csv for mappings between codes and variables.
external_data_name_to_info_list <- list(  # provide info for getting walkability data; we took a shortcut here; see the tutorial of external data for detail
        walkability=list(vector_of_expected_downloaded_file_name=NULL,
                         download_file=NULL,
                         process_file=NULL,
                         geoid_type="Block Group",
                         variable_name_to_interpolate_by_sum_boolean_mapping=walkability_variable_name_to_interpolate_by_sum_boolean_mapping

        )
)

start_time_get_variable_measures_from_acmt <- Sys.time()
city_hall_to_radius_to_variable_to_measures_list <- get_variable_measures_from_acmt(city_hall_to_lat_long_list=city_hall_to_lat_long_list, radius_vector=radius_vector, year=year, names_of_variable_to_get=names_of_variable_to_get, codes_of_acs_variables_to_get=codes_of_acs_variables_to_get, external_data_name_to_info_list=external_data_name_to_info_list)
end_time_get_variable_measures_from_acmt <- Sys.time()
```

```
print(city_hall_to_radius_to_variable_to_measures_list[1])
```

```
## $seattle_city_hall
## $seattle_city_hall$`1000`
## $seattle_city_hall$`1000`$total_pop_count
## [1] 20047.59
##
## $seattle_city_hall$`1000`$AC_UNPR
## [1] 702.4241
##
## $seattle_city_hall$`1000`$NatWalkInd
## [1] 17.8427
##
##
## $seattle_city_hall$`2000`
## $seattle_city_hall$`2000`$total_pop_count
## [1] 65734.34
##
## $seattle_city_hall$`2000`$AC_UNPR
## [1] 2419.726
##
## $seattle_city_hall$`2000`$NatWalkInd
## [1] 17.54609
##
##
## $seattle_city_hall$`3000`
## $seattle_city_hall$`3000`$total_pop_count
## [1] 124589.1
##
## $seattle_city_hall$`3000`$AC_UNPR
## [1] 5253.66
##
## $seattle_city_hall$`3000`$NatWalkInd
## [1] 17.03576
##
##
## $seattle_city_hall$`4000`
## $seattle_city_hall$`4000`$total_pop_count
## [1] 171715.4
##
## $seattle_city_hall$`4000`$AC_UNPR
## [1] 8532.688
##
## $seattle_city_hall$`4000`$NatWalkInd
## [1] 16.73851
##
##
## $seattle_city_hall$`5000`
## $seattle_city_hall$`5000`$total_pop_count
## [1] 212559.3
##
## $seattle_city_hall$`5000`$AC_UNPR
## [1] 12280.47
##
## $seattle_city_hall$`5000`$NatWalkInd
## [1] 16.51125
```

```
print("Between start and end of getting ACMT measures: ")
```

```
## [1] "Between start and end of getting ACMT measures: "
```

```
print(end_time_get_variable_measures_from_acmt - start_time_get_variable_measures_from_acmt)
```

```
## Time difference of 30.98838 mins
```

4. Calculate population density

We create a function for computing population density for the given radiues.

```
add_density_measures <- function(city_hall_to_radius_to_variable_to_measures_list) {
  city_hall_to_radius_to_variable_to_measures_with_population_density_list <- city_hall_to_radius_to_variable_to_
measures_list  # R's way of making a deep copy
  for (city_hall in names(city_hall_to_radius_to_variable_to_measures_with_population_density_list)){
    for(radius_character in names(city_hall_to_radius_to_variable_to_measures_with_population_density_list[[city_
hall]])){
      for(variable in names(city_hall_to_radius_to_variable_to_measures_with_population_density_list[[city_hall]]
[[radius_character]])) {
        if(variable == "total_pop_count"){  # add other "if" statements if you want to compute other measures
          total_pop_count <- city_hall_to_radius_to_variable_to_measures_with_population_density_list[[city_hal
l]][[radius_character]][[variable]]
          #radius_numeric <- as.numeric(radius_character)  # do not use radius
          population_density <-  total_pop_count/(city_hall_to_radius_to_variable_to_measures_with_population_den
sity_list[[city_hall]][[radius_character]][["AC_UNPR"]] * 4046.864798)
          city_hall_to_radius_to_variable_to_measures_with_population_density_list[[city_hall]][[radius_characte
r]][["population_density"]] <- population_density
        }
      }
    }
  }
  return(city_hall_to_radius_to_variable_to_measures_with_population_density_list)
}
```

```
city_hall_to_radius_to_variable_to_measures_with_population_density_list <- add_density_measures(city_hall_to_rad
ius_to_variable_to_measures_list)
```

```
print(city_hall_to_radius_to_variable_to_measures_with_population_density_list[1:2])
```

```
## $seattle_city_hall
## $seattle_city_hall$`1000`
## $seattle_city_hall$`1000`$total_pop_count
## [1] 20047.59
##
## $seattle_city_hall$`1000`$AC_UNPR
## [1] 702.4241
##
## $seattle_city_hall$`1000`$NatWalkInd
## [1] 17.8427
##
## $seattle_city_hall$`1000`$population_density
## [1] 0.007052515
##
##
## $seattle_city_hall$`2000`
## $seattle_city_hall$`2000`$total_pop_count
## [1] 65734.34
##
## $seattle_city_hall$`2000`$AC_UNPR
## [1] 2419.726
##
## $seattle_city_hall$`2000`$NatWalkInd
## [1] 17.54609
##
## $seattle_city_hall$`2000`$population_density
## [1] 0.006712857
##
##
## $seattle_city_hall$`3000`
## $seattle_city_hall$`3000`$total_pop_count
## [1] 124589.1
##
## $seattle_city_hall$`3000`$AC_UNPR
## [1] 5253.66
##
## $seattle_city_hall$`3000`$NatWalkInd
## [1] 17.03576
##
## $seattle_city_hall$`3000`$population_density
## [1] 0.005860023
##
##
## $seattle_city_hall$`4000`
## $seattle_city_hall$`4000`$total_pop_count
## [1] 171715.4
##
## $seattle_city_hall$`4000`$AC_UNPR
## [1] 8532.688
##
## $seattle_city_hall$`4000`$NatWalkInd
## [1] 16.73851
##
## $seattle_city_hall$`4000`$population_density
## [1] 0.004972842
##
##
## $seattle_city_hall$`5000`
## $seattle_city_hall$`5000`$total_pop_count
## [1] 212559.3
##
## $seattle_city_hall$`5000`$AC_UNPR
## [1] 12280.47
##
## $seattle_city_hall$`5000`$NatWalkInd
## [1] 16.51125
##
## $seattle_city_hall$`5000`$population_density
## [1] 0.004277069
##
##
##
## $los_angeles_city_hall
## $los_angeles_city_hall$`1000`
## $los_angeles_city_hall$`1000`$total_pop_count
## [1] 20247.35
##
## $los_angeles_city_hall$`1000`$AC_UNPR
```

```
## [1] 761.6894
##
## $los_angeles_city_hall$`1000`$NatWalkInd
## [1] 16.64763
##
## $los_angeles_city_hall$`1000`$population_density
## [1] 0.006568582
##
##
## $los_angeles_city_hall$`2000`
## $los_angeles_city_hall$`2000`$total_pop_count
## [1] 75624.11
##
## $los_angeles_city_hall$`2000`$AC_UNPR
## [1] 2973.994
##
## $los_angeles_city_hall$`2000`$NatWalkInd
## [1] 15.82125
##
## $los_angeles_city_hall$`2000`$population_density
## [1] 0.006283499
##
##
## $los_angeles_city_hall$`3000`
## $los_angeles_city_hall$`3000`$total_pop_count
## [1] 157247.2
##
## $los_angeles_city_hall$`3000`$AC_UNPR
## [1] 6435.669
##
## $los_angeles_city_hall$`3000`$NatWalkInd
## [1] 15.59502
##
## $los_angeles_city_hall$`3000`$population_density
## [1] 0.006037686
##
##
## $los_angeles_city_hall$`4000`
## $los_angeles_city_hall$`4000`$total_pop_count
## [1] 303410.8
##
## $los_angeles_city_hall$`4000`$AC_UNPR
## [1] 11442.5
##
## $los_angeles_city_hall$`4000`$NatWalkInd
## [1] 15.48465
##
## $los_angeles_city_hall$`4000`$population_density
## [1] 0.006552263
##
##
## $los_angeles_city_hall$`5000`
## $los_angeles_city_hall$`5000`$total_pop_count
## [1] 490979.8
##
## $los_angeles_city_hall$`5000`$AC_UNPR
## [1] 18032.81
##
## $los_angeles_city_hall$`5000`$NatWalkInd
## [1] 15.23306
##
## $los_angeles_city_hall$`5000`$population_density
## [1] 0.006727933
```

5. Plot density vs. radius, categoried by cities

We create a function for converting the data we have so far to `ggplot` friendly format. We then use `ggplot` to create our plot.

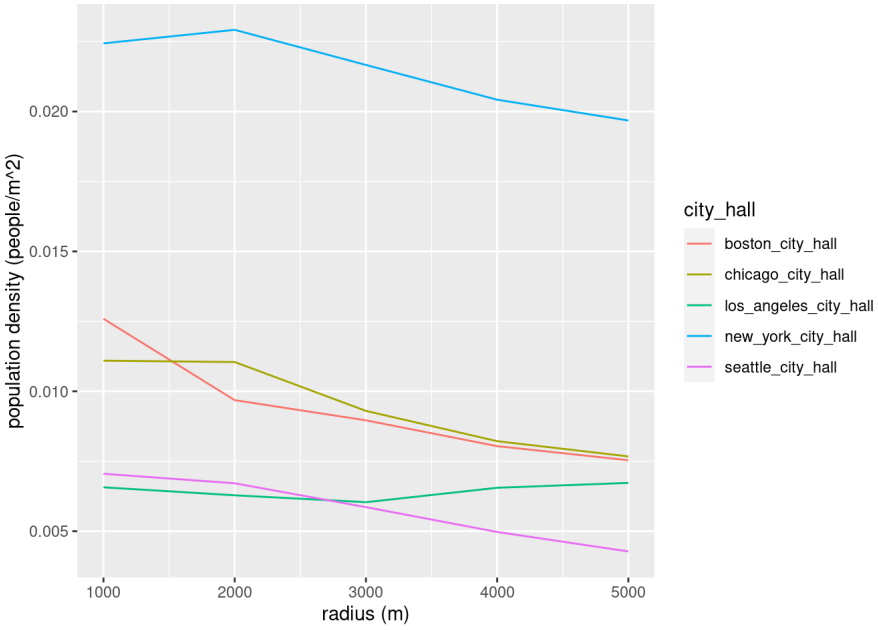
```
convert_to_dataframe_for_plotting <- function (city_hall_to_radius_to_variable_to_measures_with_population_densit
y_list) {
  city_hall_vector <- c()
  radius_vector <- c()
  variable_vector <- c()
  value_vector <- c()

  for(city_hall in names(city_hall_to_radius_to_variable_to_measures_with_population_density_list)){
    for(radius_character in names(city_hall_to_radius_to_variable_to_measures_with_population_density_list[[city_
hall]])){
      for(variable in names(city_hall_to_radius_to_variable_to_measures_with_population_density_list[[city_hall]]
[[radius_character]])){
        city_hall_vector <- c(city_hall_vector, city_hall)
        radius_vector <- c(radius_vector, as.numeric(radius_character))
        variable_vector <- c(variable_vector, variable)
        value_vector <- c(value_vector, city_hall_to_radius_to_variable_to_measures_with_population_density_list
[[city_hall]][[radius_character]][[variable]])
      }
    }
  }

  dataframe_for_plotting <- data.frame(city_hall_vector, radius_vector, variable_vector, value_vector, stringsAsF
actors=FALSE)
  names(dataframe_for_plotting) <- c("city_hall", "radius", "variable", "value")
  return(dataframe_for_plotting)
}
```

```
dataframe_for_plotting <- convert_to_dataframe_for_plotting(city_hall_to_radius_to_variable_to_measures_with_popu
lation_density_list)
```

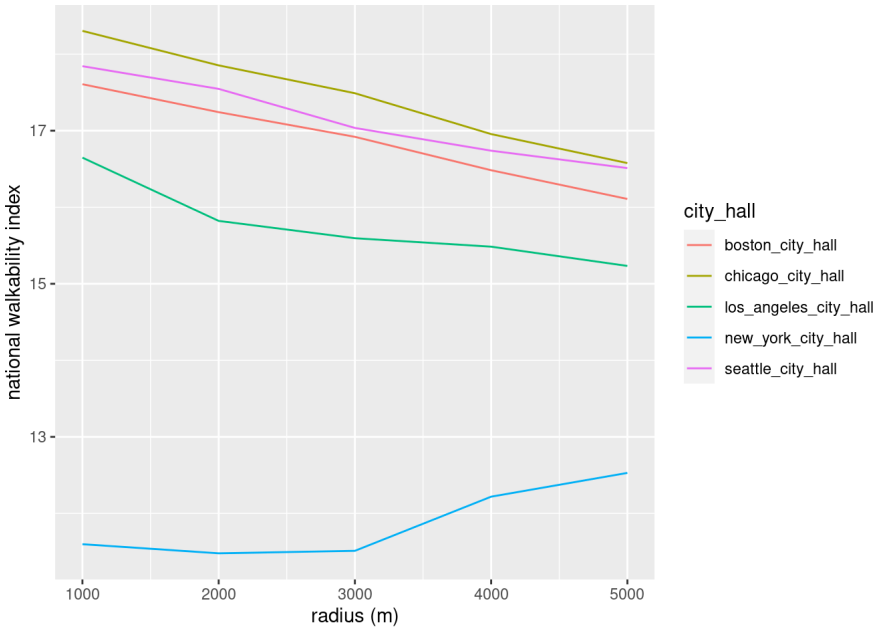
```
print(head(dataframe_for_plotting))
```

```
##           city_hall radius           variable        value
## 1 seattle_city_hall   1000    total_pop_count 2.004759e+04
## 2 seattle_city_hall   1000            AC_UNPR 7.024241e+02
## 3 seattle_city_hall   1000         NatWalkInd 1.784270e+01
## 4 seattle_city_hall   1000 population_density 7.052515e-03
## 5 seattle_city_hall   2000    total_pop_count 6.573434e+04
## 6 seattle_city_hall   2000            AC_UNPR 2.419726e+03
```

```
ggplot(dataframe_for_plotting[dataframe_for_plotting$variable == "population_density", ], aes(x=radius)) +
        geom_line(aes(y=value, col=city_hall)) +
        labs(y="population density (people/m^2)",
             x="radius (m)")
```



```
ggplot(dataframe_for_plotting[dataframe_for_plotting$variable == "NatWalkInd", ], aes(x=radius)) +
       geom_line(aes(y=value, col=city_hall)) +
       labs(y="national walkability index",
            x="radius (m)")
```



Results

We see that New York City still has a much higher population density than others, but its density reduction is less drastic than we saw when not using land area to calculate population density. Also, because of the use of land area as denominator, the population density of all cities all increased, and LA increased the least because other cities have a higher proportion of water and thus benefits more from switching from using a circular area as denominator to using land area.

Regarding to the walkability index, we see New York City is again at the top. According to the definition of walkability index, New York City has high mix of employment types and occupied housing, mix of employment types, street intersection density, and proportion of carpool workers (https://www.epa.gov/smartgrowth/smart-location-mapping# (https://www.epa.gov/smartgrowth/smart-location-mapping#):~:text=National%20Walkability%20Index%20score.,results%20of%20indicator%20rank%20scores.&text=The%20mix%20of%20employment%20typ