Creative Economy Indicator Project Technical Memo

Written by Seleeke Flingai, updated by Rachel Bowers

Date: December 27, 2022

Table of Contents:

1. Purpose of project
2. Technical details on data used
3. Project folders and R scripts
4. Example use cases of creative economy work
   a. Municipal analysis
   b. Neighborhood (block group aggregation) analysis

1. Purpose of project:

- o The New England Foundation for the Arts (NEFA) has worked to quantify impacts of the creative economy at the state/New England level
- o At municipal level, data processing, analysis, and mapping of arts and culture has been possible, but difficult
- o **Objective: can we automate creative economy data analysis and mapping at the Metro Boston, municipal, and neighborhood level?**

2. Technical details on data used:

Data source: DataAxle Massachusetts data, 2021

- o Geocoded data on over 280,000 individual establishments in Massachusetts
  - Company
  - Location (longitude, latitude)
  - NAICS 6-digit code and description
  - Employment count
- o NAICS codes: 2012

3. Project folders and R scripts

GitHub repository: https://github.com/MAPC/creative-economy

Main project folder (K drive)

The main project folder has several subfolders, a few of which will be of importance in future work. The folders of interest are:

- **Documentation** folder – contains project scope and background New England Foundation for the Arts (NEFA) reports, used to determine which NAICS codes are indicative of 'creative establishments'. Folder also includes this technical memo.
- **Data** folder – this contains the raw DataAxle data, the modified datasets that are produced by the R scripts (outlined below), and a GIS folder containing shapefiles used in the mapping analyses (all mapping is coded in R, but .csv files are produced at every crucial step that may be translated to mapping via ArcGIS).
- **R code** folder – contains all R scripts used in analyses (outlined in detail below)

- **Input** folder – contains .csv files used for neighborhood (block group aggregation) analyses.
- **Output** folder – contains the municipal level outputs of the final "Analysis" R scripts, along with regional level maps produced by the "Mapping" R script

R code folder

The R scripts folder is broken down into four subfolders – Cleaning, Filtering, Analysis, and Mapping – that were used to go from raw DataAxle 2021 master file to municipal- and regional-level summaries and maps of creative economy data.

While each script folder and its contents will be described below, it is likely that the cleaning and filtering scripts won't need to be touched for any future analysis unless updated DataAxle data is purchased or an alternative data source is considered.

1. **Cleaning** – cleans up DataAxle 2021 dataset and selects the most relevant variables for the rest of the analysis
   - Input: DataAxle 2021 data (found in Raw data folder under *data_axle_2021.csv*)
   - Output: two .csv files (found in 2021 Modified data folder):
       o *dataaxle_mass_processed.csv* (includes DataAxle data for all of Massachusetts)
       o *dataaxle_mapc_towns_processed.csv* (only includes data from MAPC municipalities)

2. **Filtering** – filters the DataAxle data to isolate creative economy establishments (per NEFA's definition). Also designates which creative category each establishment is a part of by six-digit NAICS code (creative categories also from NEFA's creative economy work).
   - Input: dataaxle_mass_processed.csv (found in Modified data folder, created through the Cleaning script)
   - Output: four .csv files (found in Modified data folder):
       o *DataAxle_nefa_all_mapc_processed.csv* (includes all creative economy establishments in MAPC)
       o *DataAxle_nefa_all_mass_processed.csv* (includes all creative economy establishments across Massachusetts)
       o *DataAxle_nefa_core_mapc_processed.csv* (includes core (as opposed to peripheral) creative economy establishments in MAPC – these are most important!)
       o *DataAxle_nefa_core_mass_processed.csv* (includes core creative economy establishments throughout Massachusetts)

3. **Analysis** – this folder contains seven scripts which summarize the data for a desired neighborhood (useful for intra-municipal analyses) or for each municipality (useful for regional analyses and comparisons between municipalities).
   a. Scripts:
       i. `01 summary stats.R` – creates summary statistics for each municipality, community type, subregion, subtype, and for MAPC and Massachusetts as a whole. Summary statistic variable list can be found in the ce summary metadata csv.

ii. `02 final dataset processed.R` — uses the `01_summary_stats.R` file (via the 'source' function in R) to create clean excel spreadsheets of the summary statistics. The "final datasets" produced from this script are then used for the final regional- and municipal-level data outputs (graphs, maps, charts)

iii. `03 final dataset percents processed.R` — same as 02-`final_dataset_processed.R`, but transforms all proportions to percentages. The "final datasets" produced from this script are also used in the final regional- and municipal-level data outputs where necessary

iv. `04 final municipal output functions processed.R` — this script builds functions to
   - Create both online (leaflet) and png dot maps and heat maps. Each dot in the dot map is a creative economy establishment, colored by NEFA creative category. The heat map illustrates the density of the creative economy establishments within a municipality.
   - Produce summary statistics excel spreadsheets and charts (outlined in greater detail below)

```
1   ##################################################################################
2   # Code: Final Municipal Output Code
3   # Purpose: This code uses the functions created in "final_municipal_output_functions.R"
4   #          to output municipality-specific graphs and a map of core creative economy establishments
5   # Author: Seleeke Flingai
6   # Date: July 23, 2018
7   ##################################################################################
8
9   # clear workspace
10  rm(list=ls())
11
12  # load in the final municipal output functions R file
13  analysis.path <- "K:/DataServices/Projects/Current_Projects/Arts_and_Culture_Planning/Creative Economy/R code/3_Analysis"
14  setwd(analysis.path)
15  source("final_municipal_output_functions_processed.R")
16
17  # Place name of municipality in quotation marks for function below
18  # output is store in Output folder of data project
19  municipal.output("Arlington")
20
```

Figure 1 -- R code from final_municipal_outputs_processed.R. Simply replace "Arlington" in the municipal.output function with desired municipality to produce municipal-specific data folder containing maps, graphs, and datasets

4. **Final Output** – This folder contains a script that produces maps and data summaries of the desired municipality (useful for municipality-specific interests) or neighborhood (as delineated by block groups that the user inputs), or produces regional maps using the summary stats data analyzed in the "Analysis" R scripts

   a. Outputs of `final_outputs_processed_all_geographies.R` (found in the Output folder)

      i. **Heat map of the core creative economy in the neighborhood/municipality** (online leaflet and png file) – *heatmap_[neighborhood/municipality]_ce.html or .png*

      ii. **Dot map of each core creative economy establishment in the neighborhood/municipality** (online leaflet and png file)) – *map_[neighborhood/municipality]_ce.html or .png* – *GENERALLY ONLY USED FOR INTERNAL PURPOSES* (map labeled to illustrate this point)

iii. **Bar graph comparing share of creative economy in desired municipality to its community type, county, subtype, subregion, and MAPC and Massachusetts as a whole** – *[municipality]_ce_core_share_comparison.png*
- Bar graph is not produced for block group/neighborhood analysis

iv. **Bar graph breaking down the number of creative establishments in the desired municipality, by NEFA creative category** (e.g., "Architecture and Design" or "Visual Arts, Music, and Other Performing Arts")
- Bar graph is not produced for block group/neighborhood analysis

v. **Bar graph breaking down the number of creative establishments in the desired municipality, by NAICS description** (e.g., "Graphic Design Services" or "Museums")
- Bar graph is not produced for block group/neighborhood analysis

vi. *The following spreadsheets:*
- **the list of creative economy establishments** – *[neighborhood/municipality]_ce_establishment_list.csv*
  - essentially a filter of the DataAxle creative economy dataset (produced from the "Filtering" R script) that only shows the creative establishments from the desired block group/neighborhood/ municipality
- **a comparison of the neighborhood/municipality's creative economy to that of its respective community type, county, subregion, subtype, as well as MAPC and Massachusetts** – [municipality]_ce_data.csv
  - for the block group/neighborhood analysis, the comparison is only to the municipality of the given block group/neighborhood
- **A breakdown of the neighborhood/municipality's creative economy by NAICS description** (e.g., "Graphic Design Services" or "Museums") – *[neighborhood/municipality]_ce_naics_desc_breakdown.csv*
  - includes number of establishments in each category and the share of that category within the neighborhood/municipality's total core creative economy
- **A breakdown of the neighborhood/municipality's creative economy by NEFA creative category** (e.g., "Architecture and Design" or "Visual Arts, Music, and Other Performing Arts") – *[neighborhood/municipality]_ce_creative_group_breakdown.csv*

4. Example use cases of Creative Economy work

Municipal analysis:

If you simply want to obtain maps, graphs, and tabular summary statistics of a given municipality: Open the final_outputs_processed_all_geographies.R script and use the
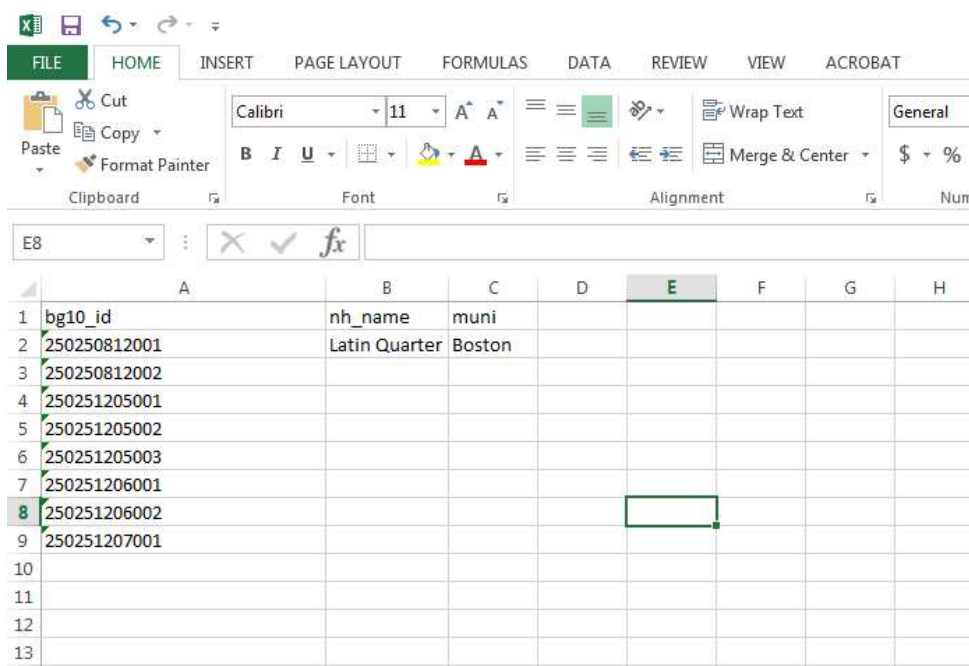
`municipal.output` function to signify which municipality's data you wish to process. The output can be found in the [Output folder](#) with the folder titled by the municipality's name.

If you want tabular data to put into ArcGIS for more customized mapping of a given municipality's creative economy: Open the [`final_municipal_outputs_processed.R`](#) script and use the `municipal.output` function to signify which municipality's data you wish to process. Open the municipality's folder in the Output folder and find the *[municipality]_ce_establishment_list.csv* creative economy establishments list. Import .csv into ArcGIS and use longitude/latitude columns to map each establishment.

Neighborhood (Block Group Aggregation) Analysis

If you want to obtain maps, graphs, and tabular summary statistics of a given neighborhood, as defined by a collection of block groups:

1. Open the [Input folder](#) and open the bg_nh_input_template.csv file.
2. Input the Census 2020 block group IDs (12-digit), the neighborhood name (only one neighborhood name), and municipality name (only one municipality name). For example, for an analysis of Boston's Latin Quarter, you would input the following:

3. Save the excel file (preferred file name format: [*yyyymmdd*]_bg_[*neighborhood name*]_[*municipality name*].csv) in the Input folder.
4. Open the [`final_outputs_processed_all_geographies.R`](#) script and run the `nhood.output` code. A dialog box should open (the file.choose function makes this happen): select the input file you created in steps 1-3. No code needs to be modified to run this script successfully. The output can be found in the [Output folder](#) of the year of analysis with the folder titled by the neighborhood's name.

If you want tabular data to put into ArcGIS for more customized mapping of a given neighborhood's creative economy: After completing steps 1-4, open the neighborhood's folder

in the Output folder and find the *[neighborhood]_ce_establishment_list.csv* creative economy establishments list. Import .csv into ArcGIS and use longitude/latitude columns to map each establishment.