

Mapping Demographic Indexes for the Casco Bay Watershed

July 27, 2022

Contents

Introduction	1
Load Libraries	2
Set Graphics Theme	2
Load Data	2
Folder References	2
Tabular Data	2
Geospatial Data	3
Merge Tabular Data and Geospatial Data	4
Initial Choropleth Map	4
Portland Region	5
A Faceted Map Display	6
Building the Tibble	6
Whole Watershed Facet Plot	6
Portland Region	8
Observed Percentiles of Indicators	9
Whole Watershed	10
Portland Region	11

Introduction

CBEP, like other National Estuary Programs will receive additional funding to support our programs via the “Bipartisan Infrastructure Law” signed into law last December.

EPA has recently released guidance for applying for those funds. A core component of the guidance is that overall, the NEP program should comply with the White House’s “Justice 40” initiative, which requires that “at least 40% of the benefits and investments from BIL funding flow to disadvantaged communities.”

EPA suggested that we use the National-scale EJSCREEN tools to help identify “disadvantaged communities” in our region. The EPA guidance goes on to suggest we focus on five demographic indicators:

- Percent low-income;
- Percent linguistically isolated;
- Percent less than high school education;
- Percent unemployed; and
- Low life expectancy.

This notebook builds on the work in “Calc_CB_Indexes.pdf” to produce simple maps showing the distribution of these five demographic metrics and composite indexes based on national-scale percentiles on each.

Load Libraries

```
library(tidyverse)
#> -- Attaching packages ----- tidyverse 1.3.1 --
#> #> v ggplot2 3.3.6     v purrrr   0.3.4
#> #> v tibble  3.1.7     v dplyr    1.0.9
#> #> v tidyverse 1.2.0    v stringr  1.4.0
#> #> v readr   2.1.2     v forcats  0.5.1
#> #> -- Conflicts ----- tidyverse_conflicts() --
#> #> x dplyr::filter() masks stats::filter()
#> #> x dplyr::lag()   masks stats::lag()
library(readr)
library(sf)      # automatically loads `sp` and `rgdal`
#> Linking to GEOS 3.9.1, GDAL 3.4.3, PROJ 7.2.1; sf_use_s2() is TRUE
```

Set Graphics Theme

This sets `ggplot()` graphics for a blank slate, suitable for maps.

```
theme_set(theme_void())
```

Load Data

Folder References

I use folder references to allow limited indirection, thus making code from GitHub repositories more likely to run “out of the box”.

```
data_folder <- "Original_Data"
gis_folder <- "GIS_Data"
dir.create(file.path(getwd(), 'figures'), showWarnings = FALSE)
```

Tabular Data

```
cb_data <- read_csv("cb_block_groups_indexes.csv",
                     col_types = paste0('ccc', rep('d', 17)))
```

Geospatial Data

We have geospatial data in UTM coordinates. Files are large, so not all are included in the github repository

```
the_file_name <- 'casco_block_groups_utm.shp'
the_path <- file.path(gis_folder, the_file_name)
cb_geospatial <- st_read(the_path )
#> Reading layer `casco_block_groups_utm` from data source
#>   `C:\Users\curtis.bohlen\Documents\Data Analysis\Casco_EJScreen\GIS_Data\casco_block_groups_utm.shp`
#>   using driver `ESRI Shapefile'
#> Simple feature collection with 231 features and 47 fields
#> Geometry type: POLYGON
#> Dimension:      XY
#> Bounding box:  xmin: 339455.5 ymin: 4820828 xmax: 438495.5 ymax: 4919775
#> Projected CRS: NAD83 / UTM zone 19N

the_file_name <- 'casco_watershed_utm.shp'
the_path <- file.path(gis_folder, the_file_name)
cb_watershed <- st_read(the_path )
#> Reading layer `casco_watershed_utm` from data source
#>   `C:\Users\curtis.bohlen\Documents\Data Analysis\Casco_EJScreen\GIS_Data\casco_watershed_utm.shp`
#>   using driver `ESRI Shapefile'
#> Simple feature collection with 817 features and 9 fields
#> Geometry type: POLYGON
#> Dimension:      XY
#> Bounding box:  xmin: 349981.6 ymin: 4823492 xmax: 433681.5 ymax: 4917055
#> Projected CRS: NAD83 / UTM zone 19N

the_file_name <- 'watershed_outline.shp'
the_path <- file.path(gis_folder, the_file_name)
cb_watershed_outline <- st_read(the_path )
#> Reading layer `watershed_outline` from data source
#>   `C:\Users\curtis.bohlen\Documents\Data Analysis\Casco_EJScreen\GIS_Data\watershed_outline.shp`
#>   using driver `ESRI Shapefile'
#> Simple feature collection with 1 feature and 1 field
#> Geometry type: LINESTRING
#> Dimension:      XY
#> Bounding box:  xmin: 349996 ymin: 4823674 xmax: 433678.3 ymax: 4917018
#> Projected CRS: NAD83 / UTM zone 19N

the_file_name <- 'Maine.shp'
the_path <- file.path(gis_folder, the_file_name)
Maine <- st_read(the_path )
#> Reading layer `Maine` from data source
#>   `C:\Users\curtis.bohlen\Documents\Data Analysis\Casco_EJScreen\GIS_Data\Maine.shp`
#>   using driver `ESRI Shapefile'
#> Simple feature collection with 1 feature and 5 fields
#> Geometry type: MULTIPOLYGON
#> Dimension:      XY
```

```
#> Bounding box: xmin: 336630.3 ymin: 4759246 xmax: 662458.3 ymax: 5256292
#> Projected CRS: NAD83 / UTM zone 19N
```

Merge Tabular Data and Geospatial Data

```
tmp <- cb_data %>%
  select(c(ID, LIFEEXP, NEG_LIFEEXP,
          Index_Raw:Index_best_Ptiles))

cb_geospatial_2<- cb_geospatial %>%
  full_join(tmp, by = 'ID')

names(cb_geospatial_2)
#> [1] "OBJECTID"           "ID"                  "MINORPCT"
#> [4] "LOWINCPCT"          "LESSHSPCT"          "LINGISOPCT"
#> [7] "UNDER5PCT"          "OVER64PCT"          "PRE1960PCT"
#> [10] "UNEMPPCT"          "DSLPM"              "CANCER"
#> [13] "RESP"               "PTRAF"              "PWDIS"
#> [16] "PNPL"               "PRMP"              "PTSDF"
#> [19] "OZONE"              "PM25"              "UST"
#> [22] "P_MINORPCT"        "P_LWINCPCT"        "P_LESHSPCT"
#> [25] "P_LNGISPCT"        "P_UNDR5PCT"        "P_OVR64PCT"
#> [28] "P_LDPNT"            "P_UNEMPPCT"        "P_DSLPM"
#> [31] "P_CANCER"           "P_RESP"             "P_PTTRAF"
#> [34] "P_PWDIS"            "P_PNPL"             "P_PRMP"
#> [37] "P PTSDF"            "P_OZONE"            "P_PM25"
#> [40] "P_UST"               "B_UST_D2"           "AREALAND"
#> [43] "AREAWATER"          "NPL_CNT"            "TSDF_CNT"
#> [46] "Shape_Leng"          "Shape_Area"         "LIFEEXP"
#> [49] "NEG_LIFEEXP"         "Index_Raw"           "Index_5_Ptiles"
#> [52] "Index_4_Ptiles"       "Index_best_Ptiles" "geometry"
```

Initial Choropleth Map

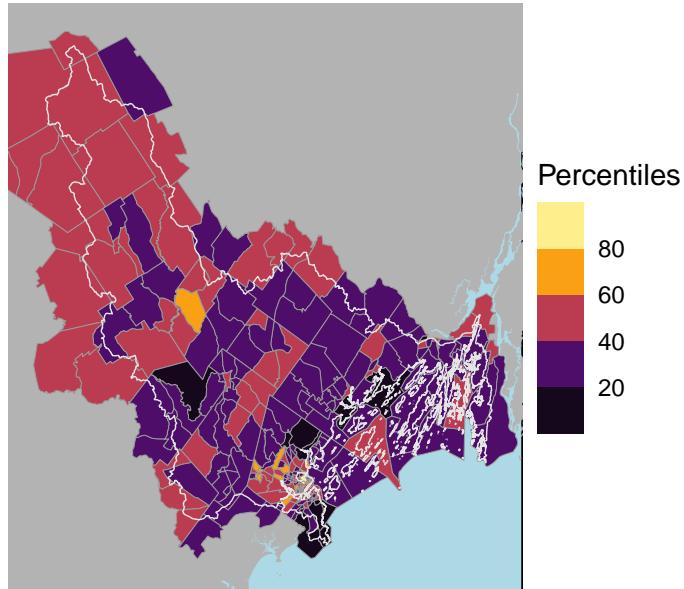
```
xlims <- c(347500, 435000)
ylims <- c(4820000, 4920000)

plt <- ggplot() +
  geom_sf(data = Maine,
          fill = 'grey70',
          color= NA) +
  geom_sf(data = cb_geospatial_2,
          aes(fill = Index_best_Ptiles),
          color="gray60",
          size = 0.1) +
  geom_sf(data = cb_watershed,
          color="grey90", fill = NA, size = 0.25) +
```

```

scale_fill_viridis_b(option = 'B', name = 'Percentiles',
                     na.value = "grey75",
                     breaks = c(0, 20, 40, 60, 80, 100)) +
  theme(panel.background = element_rect(fill = 'lightblue')) +
  coord_sf(xlim = xlims, ylim = ylims)
plt

```



```

ggsave('figures/watershed_vulnerability_index.png', type='cairo',
       width = 3.5, height = 3.5)
ggsave('figures/watershed_vulnerability_index.pdf', device = cairo_pdf,
       width = 3.5, height = 3.5)

```

Portland Region

```

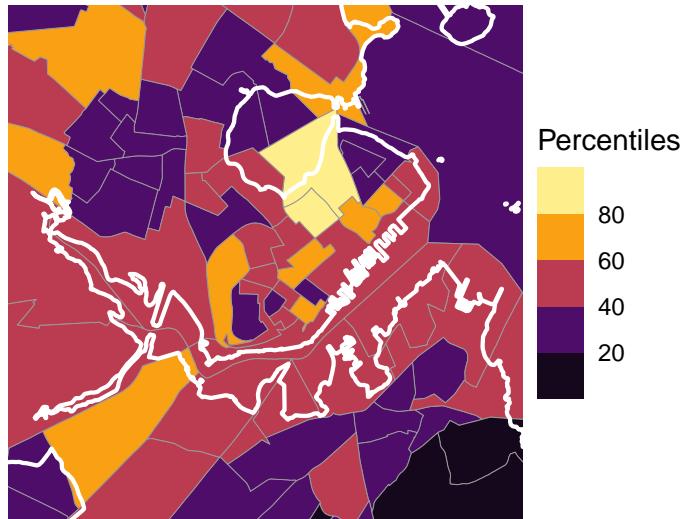
port_xlims <- c(394000, 401250)
port_ylims <- c(4830750, 4838000)

```

```

plt +
  geom_sf(data = cb_watershed, fill = NA,
           color="white", lwd = 0.75) +
  coord_sf(xlim = port_xlims, ylim = port_ylims)
#> Coordinate system already present. Adding new coordinate system, which will replace the existing one

```



```
ggsave('figures/portland_vulnerability_index.png', type='cairo',
       width = 3.5, height = 3.5)
ggsave('figures/Portland_vulnerability_index.pdf', device = cairo_pdf,
       width = 3.5, height = 3.5)
```

A Faceted Map Display

Building the Tibble

To create a `facet_wrap()` display on multiple panels, I need to build a suitable long tibble.

```
tmp <- cb_data %>%
  select(ID, P_LIFEEXP:P_UNEMPPCT) %>%
  rename(`Low Life Expectancy` = P_LIFEEXP,
        `Low Income` = P_LWINCPCT,
        `Less than High School` = P_LESHSPCT,
        `Linguistic Isolation` = P_LNGISPCT,
        `Unemployment` = P_UNEMPPCT)
```

```
cb_geospatial_3 <- cb_geospatial %>%
  left_join(tmp, by = 'ID') %>%
  pivot_longer(`Low Life Expectancy`:`Unemployment`,
             names_to = 'Threshold', values_to = 'Value')
```

Whole Watershed Facet Plot

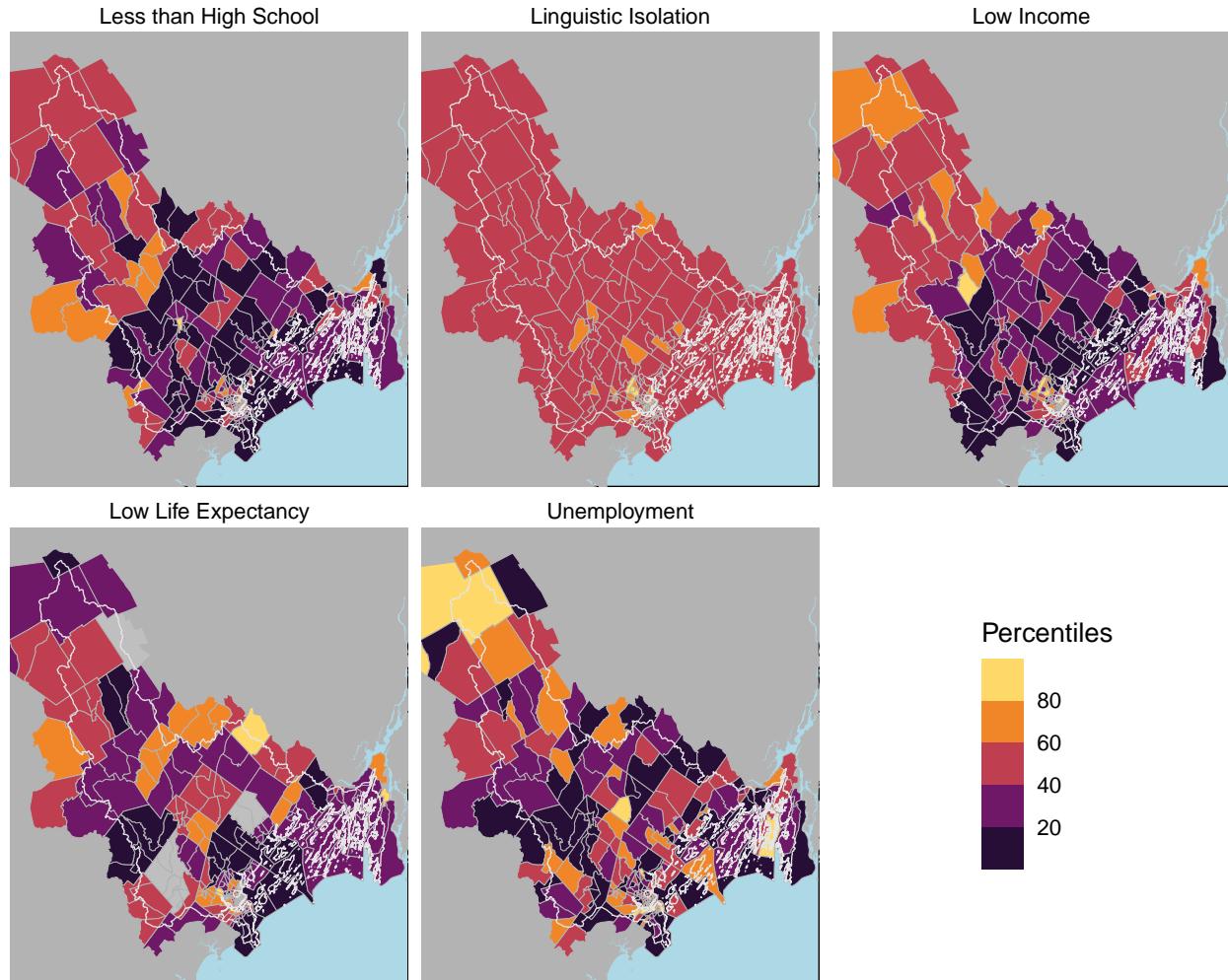
This ggplot object is fairly complex, so it takes some time to render. Rendering time can be shortened somewhat by using simpler polygon geometries.

```

plt <- ggplot() +
  geom_sf(data = Maine,
          fill = 'grey70', color=NA) +
  geom_sf(data = cb_geospatial_3, aes(fill = Value),
          color="grey70", size = 0.1) +
  facet_wrap(~ Threshold, nrow = 2) +
  scale_fill_viridis_b(option = 'B', name = 'Percentiles',
                       na.value = "grey75", breaks = c(0, 20, 40, 60, 80, 100) ) +
  theme(panel.background = element_rect(fill = 'lightblue'),
        legend.position = c(0.85, 0.25),
        strip.text.x = element_text(margin = margin(0,0,0.1,0, "cm")))
plt +
  geom_sf(data = cb_watershed, fill = NA,
          color="grey90", lwd = 0.25) +
  coord_sf(xlim = xlims, ylim = ylims) +
  ggtitle("Casco Bay Census Tracts\nNational Percentiles")

```

Casco Bay Census Tracts National Percentiles

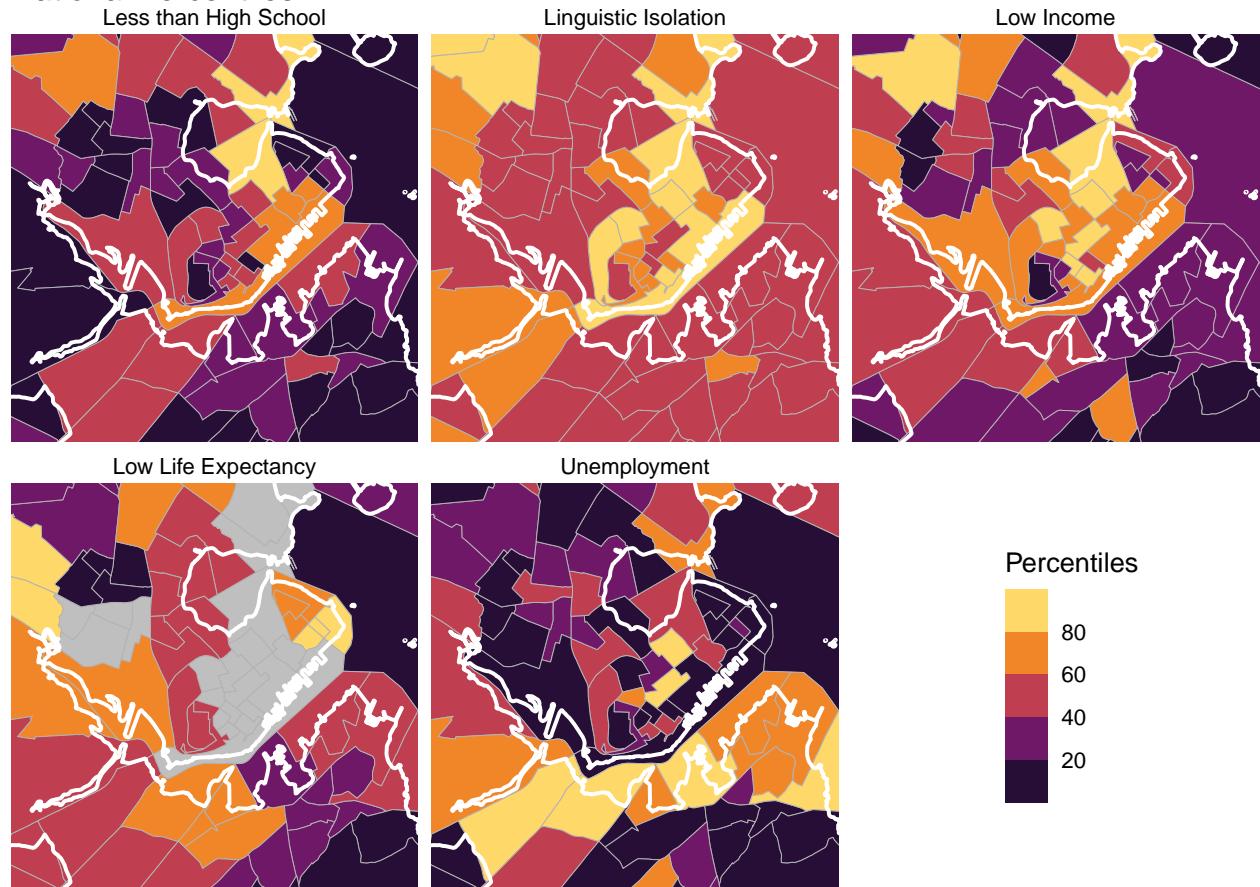


```
ggsave('figures/watershed_national.png', type='cairo',
       width = 7, height = 6)
ggsave('figures/watershed_national.pdf', device = cairo_pdf,
       width = 7, height = 6)
```

Portland Region

```
plt +
  geom_sf(data = cb_watershed, fill = NA,
           color="white", lwd = 0.75) +
  coord_sf(xlim = port_xlims, ylim = port_ylims) +
  ggtitle("Portland Region Census Tracts\nNational Percentiles")
```

Portland Region Census Tracts National Percentiles



```
ggsave('figures/portland_national.png', type='cairo',
       width = 7, height = 6)
ggsave('figures/portland_national.pdf', device = cairo_pdf,
       width = 7, height = 6)
```

So, several metrics pick up the subsidized housing and homeless community in Portland's Back Cove neighborhood. Other than that, what is striking is the fact that the ranks of the different metrics are not all that correlated in this smaller sub-region, which may help explain why few locations are flagged by any of the candidate indicators.

Observed Percentiles of Indicators

Let's look at percentiles within our region. I calculate these by calculating ranks, and dividing my the overall sample size.

Whole Watershed

```
tmp <- cb_data %>%
  select(ID, NEG_LIFEEXP, LOWINCPCT:UNEMPPCT) %>%
  mutate(across(NEG_LIFEEXP:UNEMPPCT, function(x)
               rank(x, na.last = 'keep',
                     ties.method = 'average'))/sum(! is.na(x))*100) %>%
  rename(`Low Life Expectancy` = NEG_LIFEEXP,
         `Low Income` = LOWINCPCT,
         `Less than High School` = LESSHSPCT,
         `Linguistic Isolation` = LINGISOPCT,
         `Unemployment` = UNEMPPCT)

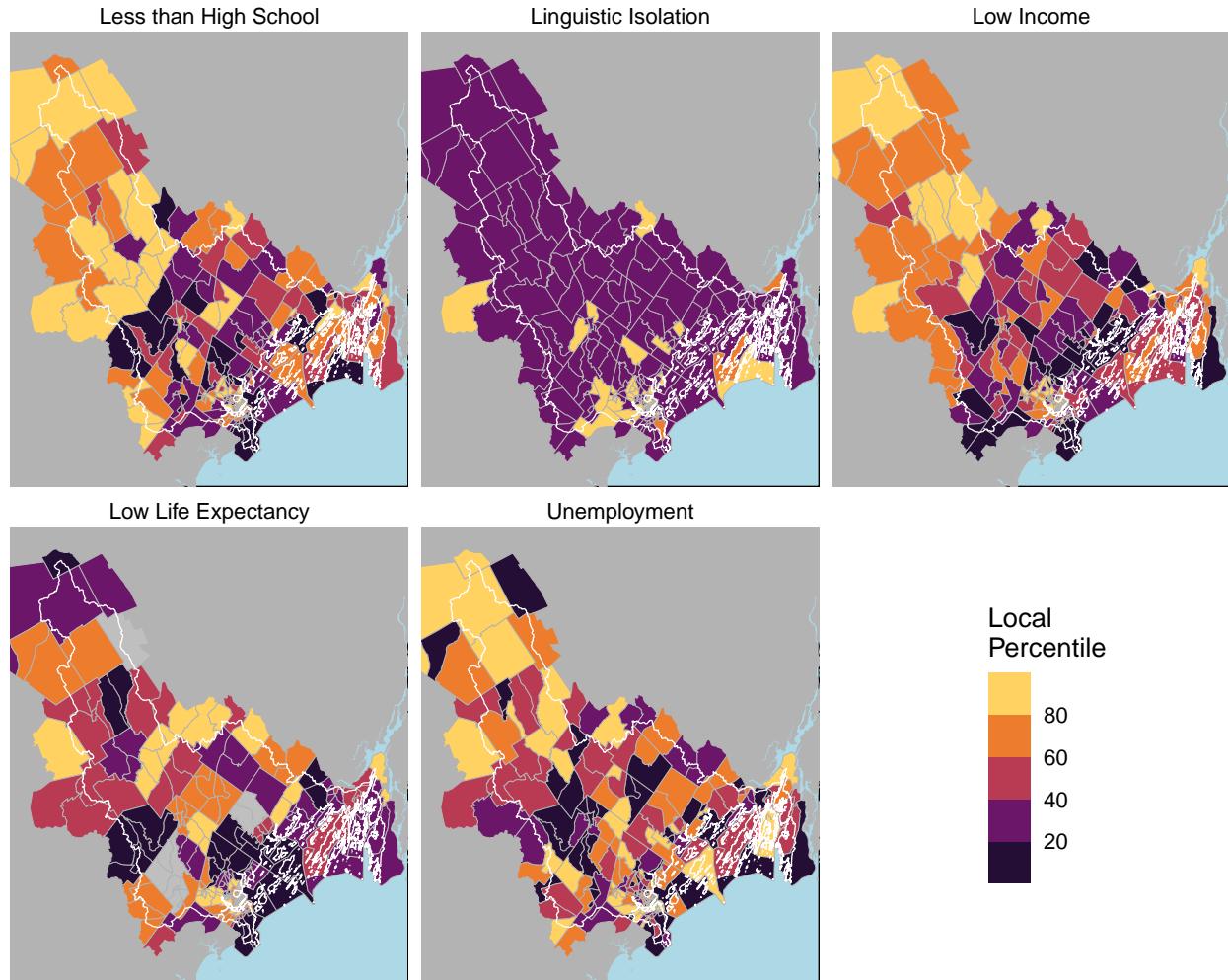
cb_geospatial_4 <- cb_geospatial %>%
  left_join(tmp, by = 'ID')

cb_geospatial_5 <- cb_geospatial_4 %>%
  pivot_longer(`Low Life Expectancy`:Unemployment,
              names_to = 'Indicator', values_to = 'Value')

plt <- ggplot() +
  geom_sf(data = Maine,
          fill = 'grey70', color=NA) +
  geom_sf(data = cb_geospatial_5, aes(fill = Value),
          color="grey70", size = 0.1) +
  facet_wrap(~ Indicator, nrow = 2) +
  scale_fill_viridis_b(option = 'B', name = 'Local\nPercentile',
                       na.value = "grey75", breaks = c(0, 20, 40, 60, 80, 100)) +
  theme(panel.background = element_rect(fill = 'lightblue'),
        legend.position = c(0.85, 0.25),
        strip.text.x = element_text(margin = margin(0,0,0.1,0, "cm")))) +
  ggtitle("Casco Bay Census Tracts\nLocal Percentiles")

plt +
  geom_sf(data = cb_watershed, fill = NA,
          color="white", lwd = 0.25) +
  coord_sf(xlim = xlims, ylim = ylims)
```

Casco Bay Census Tracts Local Percentiles



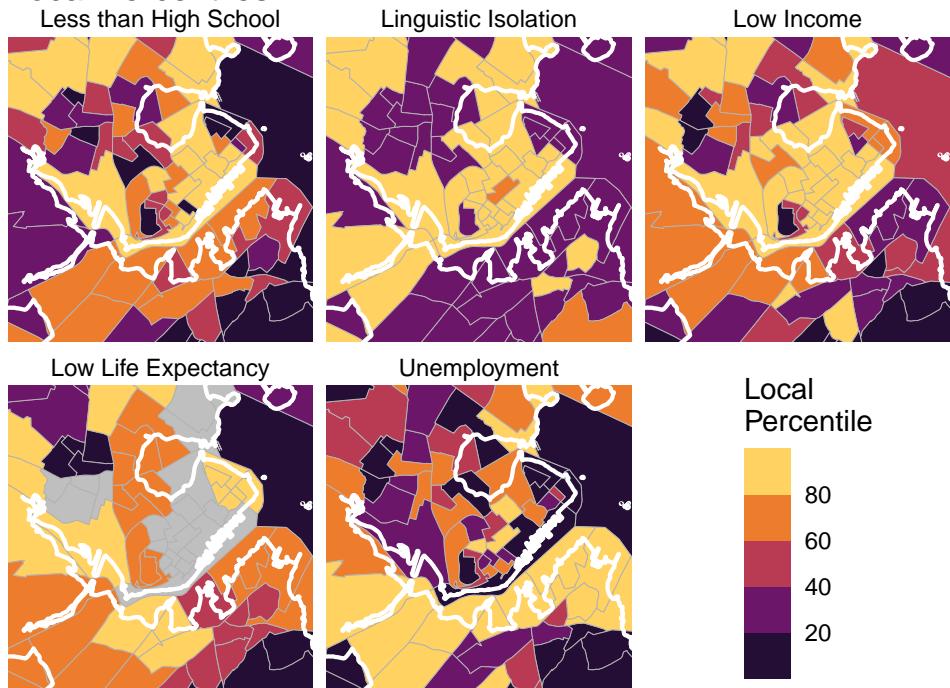
```
ggsave('figures/watershedRegional.png', type='cairo',
       width = 7, height = 6)
ggsave('figures/watershedRegional.pdf', device = cairo_pdf,
       width = 7, height = 6)
```

Portland Region

```
plt +
  geom_sf(data = cb_watershed, fill = NA,
           color="white", lwd = 0.75) +
  coord_sf(xlim = port_xlims, ylim = port_ylims) +
  ggtitle("Portland Area Census Tracts\nLocal Percentiles")
```

Portland Area Census Tracts

Local Percentiles



```
ggsave('figures/portlandRegional.png', type='cairo',
       width = 7, height = 6)
ggsave('figures/portlandRegional.pdf', device = cairo_pdf,
       width = 7, height = 6)
```