EDA Final Report Spring 2022

Reino Hyyppa Meghan Seyler

4/9/2022

Reminders for knitting commands in code chunks - can delete when we're done

- Knitting commands in code chunks:
- `include = FALSE` - code is run, but neither code nor results appear in knitted file
- `echo = FALSE` - code not included in knitted file, but results are
- `eval = FALSE` - code is not run in the knitted file
- `message = FALSE` - messages do not appear in knitted file
- `warning = FALSE` - warnings do not appear...
- `fig.cap = "..."` - adds a caption to graphical results

1. Introduction

1.1 Context

Waste diversion comes with global benefits including a decrease in
global warming potential (GWP). However, waste diversion stands to especially benefit those populations living closest to landfills by decreasing local green house gas emissions and other hazards such as odor, smoke, noise, and
water contamination.This project aims to identify which populations live closest to construction and demolition landfills in North Carolina and therefore which populations stand to benefit the most from waste diversion in the construction industry. The focus on the construction industry was inspired by one of the group member's master's project which is focused on circularity and waste diversion in the construction industry.

1.2 Research Question

Are construction and demolition landfills in North Carolina disproportionately
more common in low income counties and/or counties with high minority
populations?

1.3 Rational for Data

what's the difference between this and section 2?

2 Dataset Information

2.1 Source and Content of Data

########does this also count as table with dataset structure? Meghan askin John on 4/1

Dataset Names	Relevant Information	Source
Site Waste Facility	Lists all landfills in NC and includes landfill type and address.	NC DEQ https://deq.nc.gov/about/divisions/waste-management/solid-waste-section/solid-waste-facility-lists-presentations-and-annual-reports/solid-waste-facility-lists
CDC Social Vulnerability	Contains estimates of poverty and minority data by county.	Agency for Toxic Substances and Disease Registry https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2018.html
USA Counties Shape File	Spatial data frame that will establish the map of NC.	US Census https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html

2.2 Wrangling Process

SVI Variable Name	Variable Description	Units
E_TOTPOP	Population estimate,2014-2018 ACS	Count
E_POV	Persons below poverty estimate,2014-2018 ACS	Count
E_MINRTY	Minority estimate (all persons except white,non-Hispanic),2014-2018 ACS	Count
POV_Percent	**(E_POV/E_TOTPOP)x100**	**Percentage**
MINRTY_Percent	**(E_MINRTY/E_TOTPOP)x100**	**Percentage**

Landfill Location Variable Name	Variable Description
Address3	Landfill street name and area code
City	City name
State	State acronym
lat	**Landfill latitude**
long	**Landfill longitude**

```r
#import SVI tract-level data and change FIPS from character to factor
svi2018_nc_tract_st <-st_read('./Data/Spatial/NorthCarolinaTract/SVI2018_NORTHCAROLINA_tract.shp')
```

```
## Reading layer `SVI2018_NORTHCAROLINA_tract' from data source
##   `Z:\ENV872\ENV872-SeylerHyyppa_Final\Data\Spatial\NorthCarolinaTract\SVI2018_NORTHCAROLINA_tract.sl
##   using driver `ESRI Shapefile'
## Simple feature collection with 2192 features and 126 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -84.32187 ymin: 33.84232 xmax: -75.46062 ymax: 36.58812
## Geodetic CRS:  NAD83
```

```r
svi2018_nc_tract_st$FIPS<-as.factor(svi2018_nc_tract_st$FIPS)

#mapview(svi2018_nc_tract_st)

#import and select columns of interest from SVI county data and change FIPS from character to factor
svi2018_nc_raw <- read.csv(
  'https://svi.cdc.gov/Documents/Data/2018_SVI_Data/CSV/States_Counties/NorthCarolina_COUNTY.csv',
  colClasses = c('FIPS' = 'factor')) %>%
  select(COUNTY, FIPS, LOCATION, E_TOTPOP, E_POV, E_MINRTY)
```

3 Exploratory Analysis

(Rubric: flow between text and visualization is cohesive & relevant exploratory information is visualized)

```r
#check structure
str(svi2018_nc_raw)
```

```
## 'data.frame':    100 obs. of  6 variables:
##  $ COUNTY  : chr  "Currituck" "Dare" "Haywood" "Clay" ...
##  $ FIPS    : Factor w/ 100 levels "37001","37003",..: 27 28 44 22 90 88 15 55 85 57 ...
##  $ LOCATION: chr  "Currituck County, North Carolina" "Dare County, North Carolina" "Haywood County, M
##  $ E_TOTPOP: int  25796 35741 60433 10813 226694 33513 10447 81441 45905 34410 ...
##  $ E_POV   : int  2562 2929 8304 1700 19921 4776 1172 11593 6166 5679 ...
##  $ E_MINRTY: int  3304 4447 4300 680 62132 3249 2047 12168 4022 3684 ...
```

```r
#check colnames
colnames(svi2018_nc_raw)
```

```
## [1] "COUNTY"   "FIPS"     "LOCATION" "E_TOTPOP" "E_POV"    "E_MINRTY"
```

```r
#investigate what location column includes
glimpse(svi2018_nc_raw$LOCATION)
```

```
##  chr [1:100] "Currituck County, North Carolina" ...
```

```r
#create percent poverty and minority values by county to explore which county's have the highest percen
povdividedbytotal_col<-(svi2018_nc_raw$E_POV/svi2018_nc_raw$E_TOTPOP)*100

minrtydividedbytotal_col<-(svi2018_nc_raw$E_MINRTY/svi2018_nc_raw$E_TOTPOP)*100

#add the percentage columns to the raw data
svi2018_nc_processed<-svi2018_nc_raw%>%
  add_column(Percent_POV = povdividedbytotal_col)%>%
  add_column(Percent_MINRTY = minrtydividedbytotal_col)

svi2018_nc_poverty_top10 <- svi2018_nc_processed %>%
    arrange(desc(Percent_POV))%>%
    slice(1:10)%>%
    relocate(Percent_POV, .before = FIPS)

svi2018_nc_minrty_top10 <- svi2018_nc_processed %>%
    arrange(desc(Percent_MINRTY))%>%
    slice(1:10)%>%
   relocate(Percent_MINRTY, .before = FIPS)

#create percent poverty and minority values by tract to explore which tracts have the highest minority

povdividedbytotal_tract_col<-(svi2018_nc_tract_st$E_POV/svi2018_nc_tract_st$E_TOTPOP)*100

minrtydividedbytotal_tract_col<-(svi2018_nc_tract_st$E_MINRTY/svi2018_nc_tract_st$E_TOTPOP)*100

svi2018_nc_tract_st<-svi2018_nc_tract_st%>%
  add_column(Percent_POV = povdividedbytotal_tract_col)%>%
  add_column(Percent_MINRTY = minrtydividedbytotal_tract_col)

svi2018_nc_tract_st<-svi2018_nc_tract_st%>%
  select(COUNTY, FIPS, LOCATION, E_TOTPOP, E_POV, E_MINRTY,Percent_POV,Percent_MINRTY)

svi2018_nc_tract_st_top10 <- svi2018_nc_tract_st %>%
    arrange(desc(Percent_POV))%>%
    slice(1:10)%>%
    relocate(Percent_POV, .before = FIPS)

svi2018_nc_tract_st_top10 <- svi2018_nc_tract_st %>%
    arrange(desc(Percent_MINRTY))%>%
    slice(1:10)%>%
   relocate(Percent_MINRTY, .before = FIPS)

knitr::kable(head(svi2018_nc_poverty_top10[1:10,1:2]),
             "simple",
             col.names = c("County","Poverty Percentage"),
             align = 'c',
             caption = "Counties by Highest Poverty Percentage in North Carolina"
)
```

Table 4: Counties by Highest Poverty Percentage in North Carolina

County	Poverty Percentage
Robeson	27.35571
Jones	26.05467
Scotland	25.60830
Bladen	25.60542
Halifax	24.51630
Alleghany	24.35979

```
knitr::kable(head(svi2018_nc_minrty_top10[1:10,1:2]),
             "simple",
             col.names = c("County","Minority Percentage"),
             align = 'c',
             caption = "Counties by Highest Percentage of Minority Populations in North Carolina"
)
```

Table 5: Counties by Highest Percentage of Minority Populations
in North Carolina

County	Minority Percentage
Robeson	74.51177
Hertford	66.55488
Bertie	65.71472
Edgecombe	63.63534
Warren	61.68322
Halifax	61.46472

```
poverty_plot<-svi2018_nc_processed%>%
  arrange(Percent_POV)%>%
  mutate(COUNTY=factor(COUNTY, COUNTY))%>%
    ggplot(aes(x=COUNTY, y=Percent_POV)) +
      geom_point(color="lightgreen") +
      geom_segment( aes(x=COUNTY, xend=COUNTY, y=0, yend=Percent_POV))+
      coord_flip()+
      ggtitle("Percentage of Poverty by County in North Carolina") +
      xlab("County Name")+
      ylab("Poverty Population Percentage")+
      theme_minimal()+
      theme(
        axis.text = element_text(size = 5, face='bold', angle = 30, hjust = 1, vjust =0.5) ,
        panel.grid.major.y = element_blank(),
        panel.border = element_blank(),
        axis.ticks.y = element_blank(),
  )

poverty_plot

minority_plot<-svi2018_nc_processed%>%
  arrange(Percent_MINRTY)%>%
  mutate(COUNTY=factor(COUNTY, COUNTY))%>%
    ggplot(aes(x=COUNTY, y=Percent_MINRTY)) +
```

Figure 1: By Meghan Seyler and Reino Hyyppa

```
    geom_point(color="69b3a2") +
    geom_segment( aes(x=COUNTY, xend=COUNTY, y=0, yend=Percent_MINRTY))+
    coord_flip()+
    ggtitle("Percentage of Minority Populations by County in North Carolina") +
    xlab("County Name")+
    ylab("Minority Population Percentage")+
    theme_minimal()+
    theme(
      axis.text = element_text(size = 5, face='bold'),
      panel.grid.major.y = element_blank(),
      panel.border = element_blank(),
      axis.ticks.y = element_blank(),
      plot.subtitle = element_text("By Reino Hyyppa and Meghan Seyler"),
  )

minority_plot
```
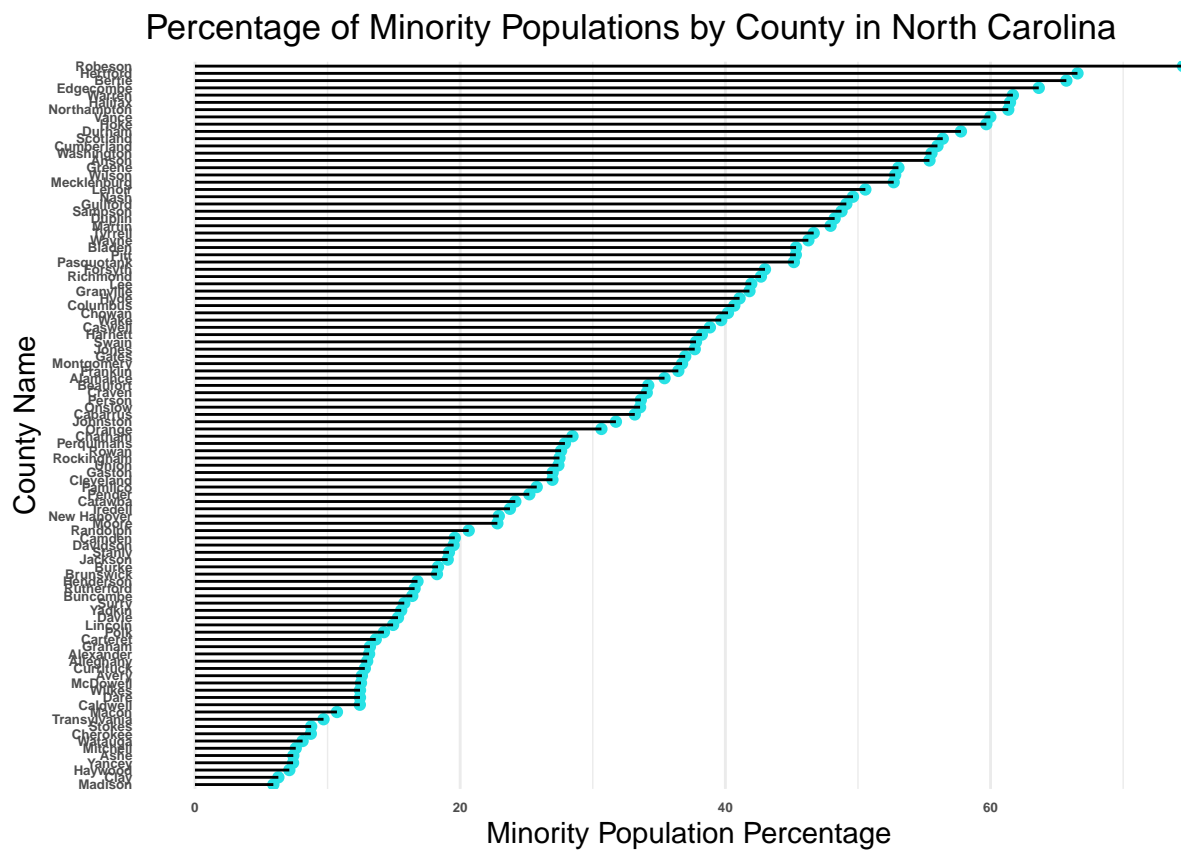

Figure 2: By Meghan Seyler and Reino Hyyppa