DC

Zhaocheng Gu

2021-03-15

Introduction

This project is one of my practice exercises for my Intro to Data Sci course, which is to use data from DC official website to plot the age distribution and car accicidents across different wards in DC, USA. The project is my own work. Please do not distribute it without my permission. The folder can also be accessed on my GitHub.

```
# Packages
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
require(ggthemes)
```

```
## Loading required package: ggthemes
```

```
require(sf)
```

```
## Loading required package: sf
## Linking to GEOS 3.10.2, GDAL 3.4.2, PROJ 8.2.1; sf_use_s2() is TRUE
```

```
require(janitor)
```

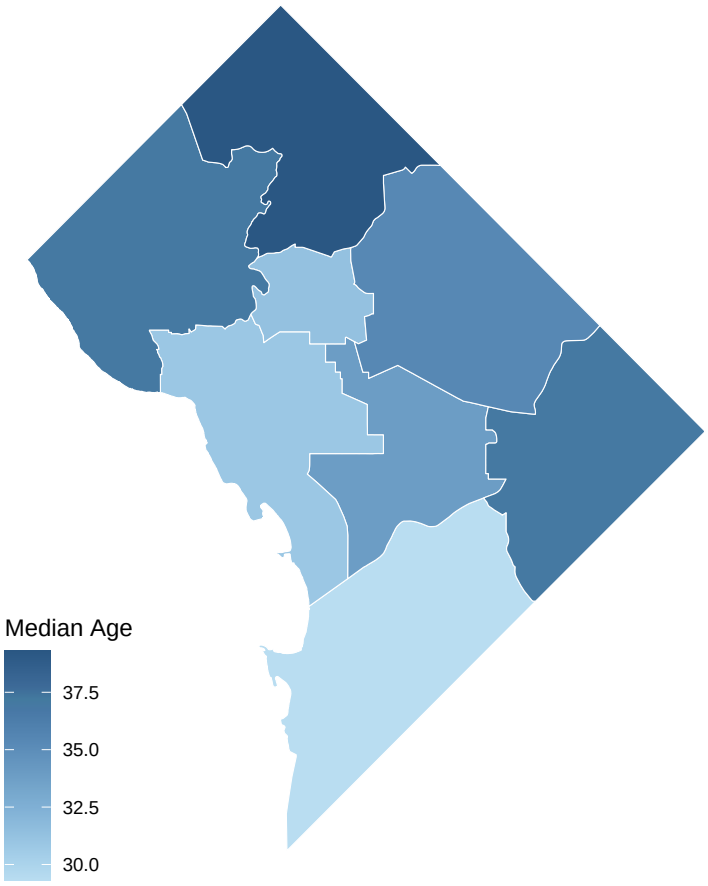
```
## Loading required package: janitor
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

Plot for the age distribution across different DC wards

Create a choropleth map that shows the median age by DC ward. Note the median_age variable can be found on the Ward_from_2012 shape data.

```
# Get the DC ward data
dc <- read_sf("dc_wards/Ward_from_2012.shp") %>%
  clean_names() %>% # clean the variable names
  mutate(median_age = as.numeric(median_age))
```

```
# Generate the choropleth map
dc %>%
  ggplot() +
  geom_sf(aes(fill=median_age),
          color = "white") +
  scale_fill_continuous_tableau() +
  labs(fill="Median Age") +
  theme_map()
```



Plot for car accidents in different DC wards

Create a choropleth map that shows the total number of car accidents in 2018 by DC Ward. Please use the map theme, provide a title for the map and legend, and use a non-default color/fill scheme.

```r
# Import the car accidents data
crashes <- read_csv("dc_crashes.csv") %>%
  clean_names() %>% # clean the variable names
  mutate(reportdate = as.Date(reportdate,"%Y/%m/%d"),
         year = lubridate::year(reportdate)) %>%
  # Convert to the time data and create a year variable
  filter(year == 2018) %>% # only use the data in 2018
  st_as_sf(coords = c("longitude", "latitude")) # convert to geometric data
```

```
## Rows: 237125 Columns: 60
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (16): CCN, REPORTDATE, ROUTEID, FROMDATE, TODATE, ADDRESS, WARD, EVENTID...
## dbl (44): X, Y, OBJECTID, CRIMEID, MEASURE, OFFSET, STREETSEGID, ROADWAYSEGI...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Ensure the crs align
st_crs(crashes) <- st_crs(dc)

# Merge the ward data with the car accidents data
crash_wards <- st_join(dc,crashes)

# Get the number of car accidents in each ward
crash_each_ward <-
  crash_wards %>%
  group_by(name) %>%
  count()

# Plot
crash_each_ward %>%
  ggplot() +
  geom_sf(aes(fill=n),color="white",alpha=.9) +
  scale_fill_viridis_c(option="magma") +
  theme_map() +
  labs(fill="Number of\nCar Accidents",
       title="Number of Car Accidents by DC Ward in 2018")
```

Number of Car Accidents by DC Ward in 2018