Rachel section

Rachel

Regression on Spatial Data

Overview of Data

Chicago has 77 neighbourhoods, which are called community areas. We can use the `read.socrata()` function to obtain the map of community areas from the Chicago Data Portal in the form of shapefiles:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(RSocrata)
library(sf)
```

```
## Linking to GEOS 3.8.0, GDAL 3.0.4, PROJ 6.3.1; sf_use_s2() is TRUE
```

```
community_map <- read.socrata("https://data.cityofchicago.org/resource/igwz-8jzy.csv")
community_map <- community_map %>% select(c(the_geom, area = area_numbe)) %>% st_as_sf(wkt = "the_geom")
```

We can also obtain a record of 3 socio-economic indicators for each of the community areas for the years 2007-2011

```
socio_ind <- read.socrata("https://data.cityofchicago.org/resource/i9hv-en6g.csv")
socio_ind <- socio_ind %>% select(-c(community_area_name))
head(socio_ind)
```

```
##   ca percent_households_below_poverty per_capita_income_ hardship_index
## 1  1                             23.6              23939             39
## 2  2                             17.2              23040             46
## 3  3                             24.0              35787             20
## 4  4                             10.9              37524             17
## 5  5                              7.5              57123              6
## 6  6                             11.4              60058              5
```

We can see there are 3 socio-economic indicators provided for each community area:

- `percent_households_below_poverty`: Percentage of households living the federal poverty line
- `per_capita_income_`: This is an estimation, calculated by aggregating incomes and dividing by the total population
- `hardship_index`: A score from 1-100 (a higher score indicates a greater level of hardship), incorporating 6 socio-economic indicators. More information on how this is calculated can be found in **REVISIT**

We now plot the map of chicago with respect to these indicators:

```
plot_map <- function(data, variable, legend.title = variable ,trans = "identity"){
  ggplot() +
    geom_sf(data = data, aes(fill = get(variable))) +
    scale_fill_viridis_c(name = legend.title,option = "magma", trans = trans) +
    theme_void()
}


library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
community_info <- left_join(community_map, socio_ind, by = c("area" = "ca"))

poverty_plot <- plot_map(data = community_info, variable = "percent_households_below_poverty",
                    legend.title = "Households \n in \n Poverty \n (%)", trans = "reverse")

income_plot <- plot_map(data = community_info, variable = "per_capita_income_",
                    legend.title = "Per Capita \n Income ($)")

hardship_plot <- plot_map(data = community_info, variable = "hardship_index",
                    legend.title = "Hardship \n Index", trans = "reverse")

grid.arrange(poverty_plot, income_plot, hardship_plot, nrow = 1)
```