

Getting started

Background

Computing the Hardship Index

Data and Social Justice: Chicago's Economic Hardship Index

The Economic Hardship Index (HI) is a useful tool to assess six indicators related to community economic development: education, employment, housing, income, dependency, and poverty. Using tables, bar graphs, and maps of HI values for Chicago's 77 community areas, the Great Cities Institute at the University of Illinois at Chicago has increased public awareness of the economic disparity between affluent neighborhoods and many neighborhoods on Chicago's South and West sides.

The HI data is based on the U.S. Census Bureau's American Community Survey, and can be accessed here (<https://greatcities.uic.edu/wp-content/uploads/2016/07/GCI-Hardship-Index-Fact-SheetV2.pdf>) for 2010-2014, and here (<https://greatcities.uic.edu/wp-content/uploads/2019/12/Hardship-Index-Fact-Sheet-2017-ACS-Final.pdf>) for the years 2013-2017. Henceforth, these date ranges will be designated '2014' (or 14) and '2017' (or 17) respectively.

Raw indicator scores and HI data for 2014 and 2017 has been compiled as an Excel file `HardshipIndex.xlsx`. We use the following abbreviations for the indicator variables:

- `Community` : name of the Chicago community area (77 in total).
- `UNEMP` : percentage over age 16 unemployed;
- `NOHS` : percentage over age 25 without a high school diploma;
- `INC` : per capita income level;
- `POV` : percentage below the federal poverty level;
- `HOUS` : percentage in housing units with more than one person per room; and
- `DEP` : percentage dependency (under 18 or over 64)

These names are further appended with a `14` or `17` to indicate the year the data refers to, e.g., `INC14` or `POV17`.

Getting started

Load packages

As before, we will explore and visualize the data using the **tidyverse** suite of packages. We also load the **readxl** library from the **tidyverse** to input our data from the Excel file.

```
library(tidyverse)
library(readxl)
```

The data

First, we'll import the Hardship Index data from an excel file, and store it in a data frame called `HI`. Type the following in your console/lab report to load the data:

```
HI <- read_excel("HardshipIndex.xlsx")
```

Use your knowledge from previous labs to answer the following questions. Feel free to go back and look at examples if you forget the correct commands.

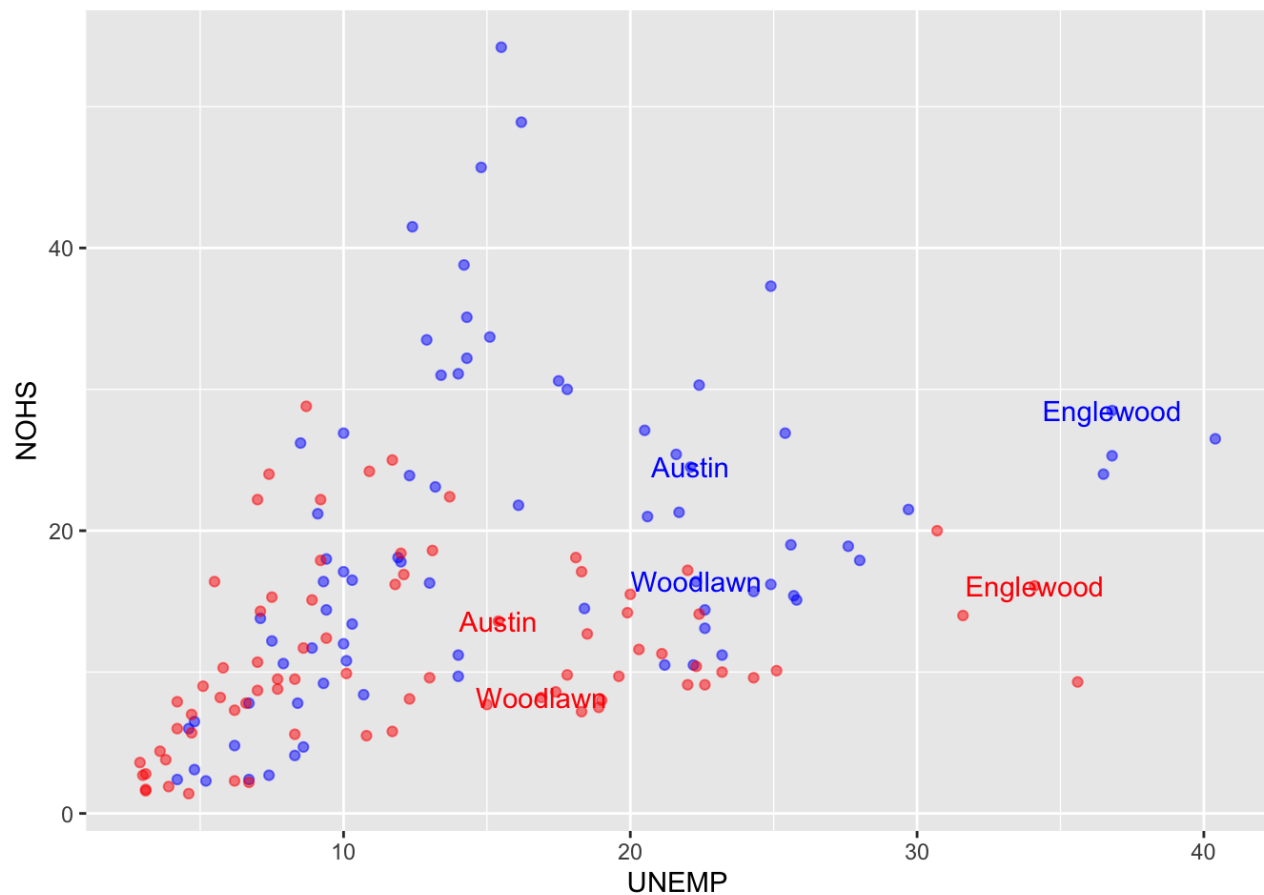
Exercise 1 What is the size of the dataframe?

Exercise 2 What command would you use to access just the columns for percentage of people living below the poverty line in 2014 and 2017? Which command would you use to access the row of scores for the Lincoln Square community?

Exercise 3 Which community had the highest per capita income in 2017? Which had the lowest?

Exercise 4 Complete the code below (fill in the `geom_????`) to get a scatter plot like the one shown, displaying unemployment rates versus education levels for both 2014 (blue) and 2017 (red). What does `alpha=.5` do? What trend do you observe for the labeled communities? What general trend do you observe overall?

```
HI_subset <- HI %>% filter(Community %in% c("Woodlawn", "Austin", "Englewood"))
ggplot(data = HI) +
  geom_????(aes(x = UNEMP14, y = NOHS14), color="blue", alpha=.5) +
  geom_????(aes(x = UNEMP17, y = NOHS17), color="red", alpha=.5) +
  geom_text(data = HI_subset, aes(x = UNEMP14, y = NOHS14, label=Community), color="blue") +
  geom_text(data = HI_subset, aes(x = UNEMP17, y = NOHS17, label=Community), color="red") +
  labs(x = "UNEMP", y = "NOHS")
```



Background

There is a long history of racism and injustice in multiple forms contributing to the current hardship experienced by many communities on the South and West sides of Chicago, e.g., redlining, slum-landlordism, white flight, mass-incarceration, etc. You can find out more about these practices from the Chicago Public Library website (<https://www.chipublib.org/housing/>) or by watching this video (<https://www.facebook.com/watch/?v=2092526767437464&extid=CGJx7og1Dlu8dgJB>), among other places.

Exercise 5

Give a brief summary of one of the forms of racism and/or injustice mentioned above (or choose your own). What difference does knowledge of a long series of historical injustices and inequities make when considering the current plight of poor community areas?

Computing the Hardship Index

The following steps are used to compute the hardship index:

- STEP ONE:** First normalize on a scale from 0 (least hardship) to 1 (greatest hardship) the raw data for each of the 6 indicators UNEMP, NOHS, DEP, HOUS, POV, and INC. The normalized value V_{NORM} of a raw indicator value V_{RAW} is obtained using $V_{NORM} = (V_{RAW} - V_{MIN}) / (V_{MAX} - V_{MIN})$, where V_{MAX} is the highest raw value for V_{RAW} and V_{MIN} is the lowest. (Note that for $V_{RAW} = INC$, the formula is $V_{NORM} = (V_{MAX} - V_{RAW}) / (V_{MAX} - V_{MIN})$ since greater income represents reduced hardship.)

- **STEP TWO:** Multiply by 100 each of the normalized indicator scores to create a 0 to 100 scale for each variable.
- **STEP THREE:** Take the average of the 6 normalized and scaled scores to get the hardship index.

There are a few ways we could go about computing the scaled scores from steps 1 and 2. One simple way would be to use the `mutate` function. This bit of code will create a new column in `HI` called `HOUS14_scaled`, and store it in the `HI` dataframe.

```
HI <- HI %>% mutate(HOUS14_scaled = 100*(HOUS14 - min(HOUS14))/(max(HOUS14)-min(HOUS14)))
```

We could then repeat this for all 6 variables. A more advanced way is given below.

Exercise 6 The code below adds columns containing the scaled scores for 2014 to the `HI` dataframe. Modify it to find the scaled scores for 2017. Then use the `mutate` function to create a new column called `HI17_computed` that is the average the scaled scores from 2017 (**Step 3** above). Show that `HI17` and `HI17_computed` are the same.

```
HI <- HI %>%
  mutate(
    across(
      HOUS14:DEP14,                                # Select all the columns of 2014 data (except income)
      ~100*(.x - min(.x)) / (max(.x) - min(.x)),    # Formula we apply to each column, ".x" is the raw score
      .names = "{col}_scaled"                        # Name each new column, for example, "HOUS14_scaled"
    ),
    # Income is different, since higher income corresponds to less economic hardship
    INC14_scaled = 100*( max(INC14) - INC14 ) / ( max(INC14) - min(INC14) )
  )
```

Normalization and Standardization

Notice that we use normalized scores in computing the Hardship Index.

Exercise 7 Reuse the code from Exercise 4 to make a scatterplot of `DEP` vs `INC`, with labels for **Austin**, **Englewood** and **Woodlawn**.

Exercise 8 The American Community Survey is constructed by combining data across a 5 year period, and are neither averages of individual years nor necessarily representative of the middle or end years. Why do you think the US Census Bureau advises against comparing 2014 and 2017 data?

Exercise 9 The method used to scale the Hardship Index data is sometimes called **min-max scaling**. Another way to scale data is with the **z-score** we covered in class, a method known as **standardization**. What is the main difference between the two methods? Why do you think the creators of the Hardship Index chose to use min-max scaling?

Thinking further about the Hardship Index

Exercise 10 Gentrification is one example of a societal injustice impacting disadvantaged communities. What major project in Woodlawn is a source of concern in regards to gentrification? Think about the different factors going into the score: what will likely happen to Woodlawn's hardship index if major gentrification occurs? Will the plight of the poorest who remain in Woodlawn be indicated by the change in hardship index?
