

Homework 1: Reproducible Computing

Ruilong Chen

2026-01-19

Dataset and Research Question

This report uses the CDC Social Vulnerability Index (SVI) dataset, which summarizes socioeconomic and demographic factors related to community resilience and vulnerability in public health contexts. The dataset is publicly available and stored in the `data/` directory of this project.

The research question addressed in this analysis is:

How is overall social vulnerability associated with the weighted average quintile and population size?

Data Cleaning

The data cleaning process is implemented in a separate R script to promote modularity and reproducibility. This step reads the raw data using relative file paths, selects relevant variables, removes observations with missing values, and saves a tidy dataset for downstream analysis.

```
source("R/01_load_clean.R")
```

Statistical Modeling

A linear regression model is fitted to assess the association between overall social vulnerability and the weighted average quintile, adjusting for total population size.

```
source("R/02_model.R")
```

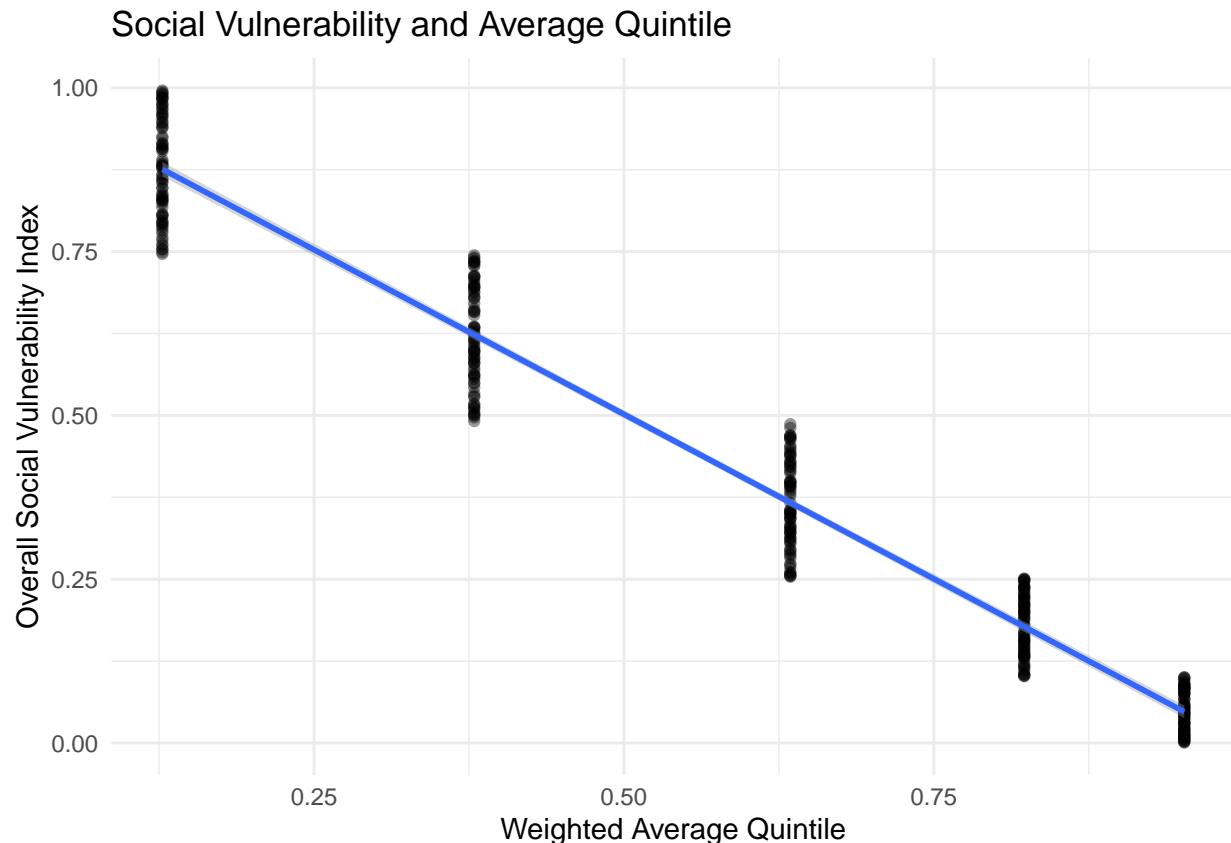
```
##  
## Call:  
## lm(formula = RPL_Themes ~ WeightedAvgQuintile + TotalPopulation,  
##      data = svi_tidy)  
##  
## Residuals:  
##       Min        1Q     Median        3Q       Max  
## -0.1266777 -0.041436 -0.000343  0.044541  0.124089  
##  
## Coefficients:  
##                               Estimate Std. Error t value Pr(>|t|)  
## (Intercept)           1.017e+00  1.180e-02  86.232   <2e-16 ***  
## WeightedAvgQuintile -1.005e+00  1.027e-02 -97.925   <2e-16 ***  
## TotalPopulation      -2.433e-06  1.768e-06  -1.376    0.17  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.06103 on 394 degrees of freedom  
## Multiple R-squared:  0.9606, Adjusted R-squared:  0.9604
```

```
## F-statistic: 4800 on 2 and 394 DF, p-value: < 2.2e-16
```

Visualization

The relationship between weighted average quintile and overall social vulnerability is visualized using a scatter plot with a fitted linear regression line to illustrate the overall trend.

```
source("R/03_visualize.R")
```



Summary of Findings

The analysis reveals a strong inverse association between weighted average quintile and the overall social vulnerability index. Areas with higher weighted average quintiles tend to exhibit substantially lower levels of social vulnerability. Because the weighted average quintile measure is closely related to the construction of the SVI, these results should be interpreted as descriptive rather than causal.

Reproducibility

```
sessionInfo()
```

```
## R version 4.3.2 (2023-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 11 x64 (build 26200)
##
## Matrix products: default
##
##
```

```

## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics   grDevices utils      datasets   methods    base
##
## other attached packages:
## [1] lubridate_1.9.4forcats_1.0.0  stringr_1.5.1  dplyr_1.1.4
## [5] purrrr_1.0.2     readr_2.1.4    tidyrr_1.3.0  tibble_3.2.1
## [9] ggplot2_3.5.1   tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] utf8_1.2.4        generics_0.1.3    stringi_1.8.3    lattice_0.21-9
## [5] hms_1.1.3         digest_0.6.33    magrittr_2.0.3   evaluate_0.23
## [9] grid_4.3.2        timechange_0.3.0 fastmap_1.1.1   Matrix_1.6-5
## [13] tinytex_0.49      mgcv_1.9-0       fansi_1.0.6     scales_1.3.0
## [17] cli_3.6.2         rlang_1.1.5      crayon_1.5.2   bit64_4.0.5
## [21] munsell_0.5.0     splines_4.3.2   withr_2.5.2     yaml_2.3.10
## [25] tools_4.3.2       parallel_4.3.2 tzdb_0.4.0      colorspace_2.1-0
## [29] vctrs_0.6.5       R6_2.5.1        lifecycle_1.0.4 bit_4.0.5
## [33] vroom_1.6.5       pkgconfig_2.0.3 pillar_1.9.0    gtable_0.3.4
## [37] glue_1.8.0         highr_0.10      xfun_0.41      tidyselect_1.2.1
## [41] rstudioapi_0.15.0 knitr_1.45     farver_2.1.1   htmltools_0.5.8.1
## [45] nlme_3.1-163      labeling_0.4.3  rmarkdown_2.29 compiler_4.3.2

```