



Introduction to Infographics and Data Visualization (Spring 2023)

20231 - JMM622-O

5th Assignment, Version 1

Yuri Souza

Project One: Homelessness

Yuri Souza

2023-03-07

This file contains the workflow I used for the project topic Homelessness

I choose to work with homeless students in New York state. I picked up this topic because this is probably the social class most affected by homelessness, given that they are young and must make decisions that will guide and impact adult life. Everything I used here is also available on my *GitHub*.

The infochartic was designed mainly in four sections. **1.** The first section includes a chart containing the total number of homeless students from 2009 to 2021. I decided to use a line chart because it is easier to see trends in time. Aligned with this chart, I included a map of students' homeless per county in New York state. I decided to use counties because I want to show that this problem is happening across the whole state.

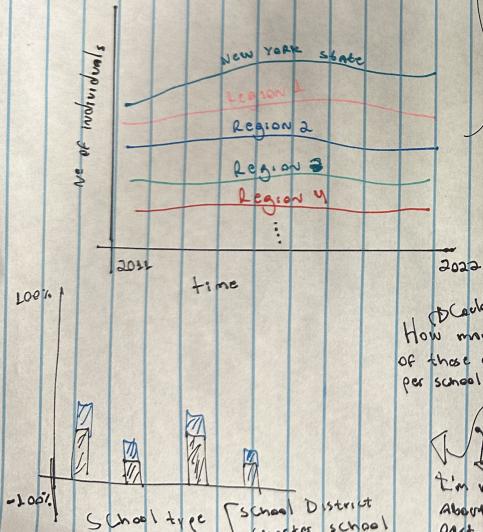
2. Given that New York has 62 counties and it would be hard to make a chart representing each of them well, I tried to show these changes in the second section using the New York economic region's boundaries. I chose a stream chart to display the data over time because I wanted to show trends in the increased proportion between the areas simultaneously. Besides, it allowed me to avoid lines overlapping within the chart. I included a map showing the proportion increase in 2021 compared to 2009 per economic region. My idea of plotting both chart and map is to show that the rise in homeless students is unrelated to the region area size and that only looking at the percentage increase does not represent the number of students per region well.

3. In the third section, I use a line chart to show which schools these students attend and how disproportional it is over time. I also included a bar chart showing the proportion changes in 2021 compared to 2009.

4. The fourth section is dedicated to the show the students' proportion per school grade over time, using a bumpchart. My choice for a bumpchat was because it allowed me to show the overlap between lines and trends better than the line chart. This section also has a bar chart comparing 2021 to 2009 for each grade. Since I do not have the data relative to students' age, I considered using grades to proxy how old these people living in homeless conditions are.

The figure below is a free-hand draft I made before designing the infographic using Adobe Illustrator®.

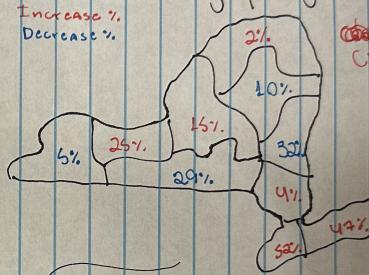
Student Homelessness
Status in New York ~~Regions~~
over the last 10 years



How much did it change per Regions?

Increase %.
Decrease %.

Choropleth map?

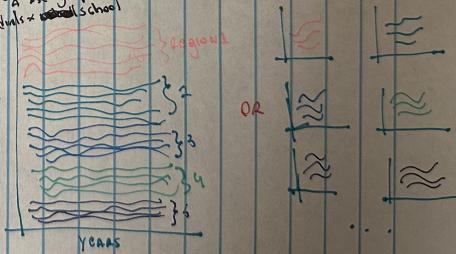


What about counties per Region?

Problem of having 62 Counties.

(Could also think this as a lot of individuals x school)

Counties



I'm not sure about this part.

Years

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

Loading the packages

```
if(!require(tidyverse))install.packages("tidyverse", dependencies = TRUE)
if(!require(readxl)) install.packages("readxl", dependencies = TRUE)
if(!require(purrr)) install.packages("purrr", dependencies = TRUE)
if(!require(writexl)) install.packages("writexl", dependencies = TRUE)
if(!require(openxlsx)) install.packages("openxlsx", dependencies = TRUE)
if(!require(data.table)) install.packages("data.table", dependencies = TRUE)
```

Setting directories and reading tables

```
path <- "C:/Users/Yuri/My Drive/PhD/00_UM/03_graduate/00_courses/02_2nd_semester/02_infographic_I/01_cl
path

regions <- readr::read_csv("C:/Users/Yuri/My Drive/PhD/00_UM/03_graduate/00_courses/02_2nd_semester/02_
unique()
regions

file_path <- list.files(path, pattern="\\.xlsx$", full.names = TRUE)
file_path

#For each file in file_path
homeless.tables <- purrr::map(file_path, ~
  #For each sheet
  purrr::map(2:3, function(i) {
    #Read the file with particular sheet number
    openxlsx::read.xlsx(.x, sheet=i, startRow=1)}) %>%
  purrr::reduce(dplyr::full_join) %>%
  #Remove all NA rows
  dplyr::filter(Reduce(`/`, across(.fns = ~!is.na(.)))) %>%
  #Add Year column at 1st position
  dplyr::mutate(YEAR = tools::file_path_sans_ext(basename(.x)), .before = 1))

homeless.tables
```

Binding the tables in a single one

```
homeless.tables.bind <- data.table::rbindlist(homeless.tables,
  use.names=TRUE,
  fill=FALSE,
  idcol=TRUE) %>%
dplyr::rename_all(., .funs = toupper) %>%
dplyr::filter(!"TOTAL.HOMELESS" == "S") %>%
dplyr::mutate(TOTAL.HOMELESS = as.numeric(TOTAL.HOMELESS),
  YEAR = as.numeric(stringr::str_extract(YEAR, ".0..")))) %>%
dplyr::left_join(regions) %>%
stats::na.omit() %>%
dplyr::select(c(3,4,6,5,2))
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```
homeless.tables.bind
```

```
#write_csv(homeless.tables.bind, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/02_2nd_seme
```

Summarizing the tables by total individuals per year

```
homeless.tables.bind.year <- homeless.tables.bind %>%
  dplyr::group_by(YEAR) %>%
  dplyr::summarise(TOTAL.HOMELESS = sum(TOTAL.HOMELESS))
homeless.tables.bind.year
```

```
#write_csv(homeless.tables.bind.year, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/02_2nd_s
```

```
homeless.tables.bind.year.diff <- homeless.tables.bind.year %>%
  dplyr::filter(YEAR == 2009 | YEAR == 2021) %>%
  tidyr::pivot_wider(names_from = YEAR, values_from = TOTAL.HOMELESS) %>%
  dplyr::mutate(DIFF = `2021` - `2009`,
                PCT_INCREASE = ((DIFF*100)/(`2009`))) %>%
  dplyr::filter(PCT_INCREASE != 0) %>%
  dplyr::filter_all(all_vars(!is.infinite(.)))
homeless.tables.bind.year.diff
```

```
#write_csv(homeless.tables.bind.year.diff, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/02_
```

Summarizing the tables by total individuals per year and region

```
homeless.tables.bind.year.region <- homeless.tables.bind %>%
  dplyr::group_by(REGION, YEAR) %>%
  dplyr::summarise(TOTAL.HOMELESS = sum(TOTAL.HOMELESS)) %>%
  dplyr::ungroup()
homeless.tables.bind.year.region
```

```
#write_csv(homeless.tables.bind.year.region, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/0
```

```
homeless.tables.bind.year.region.diff <- homeless.tables.bind.year.region %>%
  dplyr::filter(YEAR == 2009 | YEAR == 2021) %>%
  tidyr::pivot_wider(names_from = YEAR, values_from = TOTAL.HOMELESS) %>%
  dplyr::mutate(DIFF = `2021` - `2009`,
                PCT_INCREASE = ((DIFF*100)/(`2009`))) %>%
  dplyr::filter(PCT_INCREASE != 0) %>%
  dplyr::filter_all(all_vars(!is.infinite(.))) %>%
  tidyr::pivot_longer(cols = 2:3, values_to = "INDI", names_to = "YEAR") %>%
  dplyr::select(c(1,4,5,2,3))
```

```
homeless.tables.bind.year.region.diff
```

```
#write_csv(homeless.tables.bind.year.region.diff, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_cour
```

```

homeless.tables.bind.year.region.pct.diff <- homeless.tables.bind.year.region %>%
  #dplyr::filter(YEAR == 2009 | YEAR == 2021) %>%
  tidyr::pivot_wider(names_from = YEAR, values_from = TOTAL.HOMELESS) %>%
  dplyr::mutate(`2011i` = (((`2009` - `2011`)*100)/(`2009`)),
    `2012i` = (((`2011` - `2012`)*100)/(`2011`)),
    `2013i` = (((`2012` - `2013`)*100)/(`2012`)),
    `2014i` = (((`2013` - `2014`)*100)/(`2013`)),
    `2015i` = (((`2014` - `2015`)*100)/(`2014`)),
    `2016i` = (((`2015` - `2016`)*100)/(`2015`)),
    `2017i` = (((`2016` - `2017`)*100)/(`2016`)),
    `2018i` = (((`2017` - `2018`)*100)/(`2017`)),
    `2019i` = (((`2018` - `2019`)*100)/(`2018`)),
    `2020i` = (((`2019` - `2020`)*100)/(`2019`)),
    `2021i` = (((`2020` - `2021`)*100)/(`2020`))) %>%
  dplyr::filter_all(all_vars(!is.infinite(.))) #%%>
  #tidyr::pivot_longer(cols = 2:13, values_to = "DIFF_PCT", names_to = "YEAR") %>%
  #dplyr::select(c(1,3,2)) %>%
  #mutate(DIFF_PCT = round(DIFF_PCT, digits = 0),
  #       DIFF_STS = ifelse(DIFF_PCT >= 0, "Positive", "Negative"))

homeless.tables.bind.year.region.pct.diff

#write_csv(homeless.tables.bind.year.region.pct.diff, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_"

```

Summarizing the tables by total individuals per year and county

```

homeless.tables.bind.year.county <- homeless.tables.bind %>%
  dplyr::group_by(COUNTY, YEAR) %>%
  dplyr::summarise(TOTAL.HOMELESS = sum(TOTAL.HOMELESS))
homeless.tables.bind.year.county

#write_csv(homeless.tables.bind.year.county, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/00_"

homeless.tables.bind.year.county.diff <- homeless.tables.bind.year.county %>%
  dplyr::filter(YEAR == 2009 | YEAR == 2021) %>%
  tidyr::pivot_wider(names_from = YEAR, values_from = TOTAL.HOMELESS) %>%
  dplyr::mutate(DIFF = `2021` - `2009`,
    PCT_INCREASE = ((DIFF*100)/(`2009`))) %>%
  dplyr::filter(PCT_INCREASE != 0) %>%
  dplyr::filter_all(all_vars(!is.infinite(.))) %>%
  tidyr::pivot_longer(cols = 2:3, values_to = "INDI", names_to = "YEAR") %>%
  dplyr::select(c(1,4,5,2,3))
homeless.tables.bind.year.county.diff

#write_csv(homeless.tables.bind.year.county.diff, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_"

```

Summarizing the tables by total individuals per year and school type

```
homeless.tables.bind.year.school <- homeless.tables.bind %>%
  dplyr::group_by(LEA.TYPE, YEAR) %>%
  dplyr::summarise(TOTAL.HOMELESS = sum(TOTAL.HOMELESS))
homeless.tables.bind.year.school

#write_csv(homeless.tables.bind.year.school, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/02_2nd_sch

homeless.tables.bind.year.school.diff <- homeless.tables.bind.year.school %>%
  dplyr::filter(YEAR == 2009 | YEAR == 2021) %>%
  tidyr::pivot_wider(names_from = YEAR, values_from = TOTAL.HOMELESS) %>%
  dplyr::mutate(DIFF = `2021` - `2009`,
                PCT_INCREASE = ((DIFF*100)/(`2009`))) %>%
  dplyr::filter(PCT_INCREASE != 0) %>%
  dplyr::filter_all(all_vars(!is.infinite(.))) %>%
  tidyr::pivot_longer(cols = 2:3, values_to = "INDI", names_to = "YEAR") %>%
  dplyr::select(c(1,4,5,2,3))
homeless.tables.bind.year.school.diff

#write_csv(homeless.tables.bind.year.school.diff, "C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/02_2nd_sch
```

Summarizing the tables by grades

```
homeless.ny_regions <- readr::read_csv("C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/02_2nd_sch
unique()

homeless.ny_grades_regions <- readr::read_csv("C:/Users/Yuri/My Drive/PhD/00 UM/03_graduate/00_courses/02_2nd_sch
dplyr::left_join(homeless.ny_regions) %>%
dplyr::select(c(20, 3:14,19)) %>%
tidyr::pivot_longer(cols = 2:13, values_to = "Grades", names_to = "Age") %>%
#readr::write_csv("00_data/ny_homeless_student_grades_region.csv") %>%
dplyr::group_by(Year, Age) %>%
summarise(value = sum(Grades)) %>%
dplyr::ungroup() %>%
dplyr::filter(Year == "2009" | Year == "2021") %>%
tidyr::pivot_wider(names_from = "Year", values_from = "value") %>%
dplyr::mutate(DIFF = `2021` - `2009`,
                PCT_INCREASE = ((DIFF*100)/(`2009`))) %>%
dplyr::filter(PCT_INCREASE != 0) #%>%
#readr::write_csv("00_data/ny_homeless_student_grades_region_diff.csv")
```

First Version

The page below is the first version of the project proposed here. Some considerations I have for the second version:

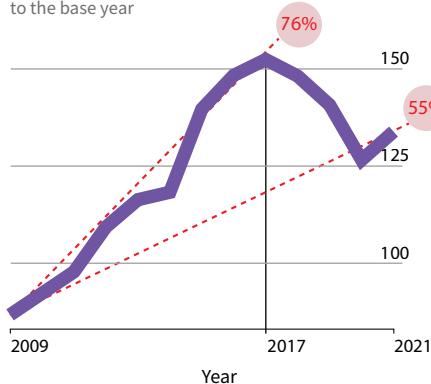
1. Maybe remove the bar chart of section three or include it within the line chart in this section since this bar chart seems a little redundant.
2. Finding some neat way to replace or show the bumpchat chart in section four.
3. Draw an image at the top of the last chart of section four.
4. Include information about where these students live when they are not in school.

Graphic detail Homeless students in New York

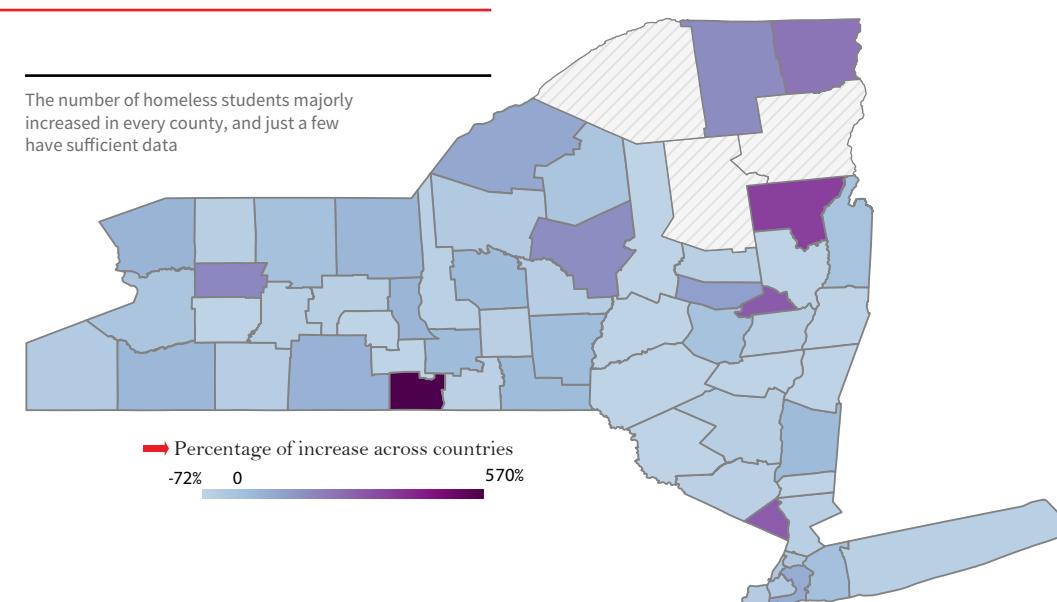
The Economist March 07, 2023

New York homelessness students increased over the last decade

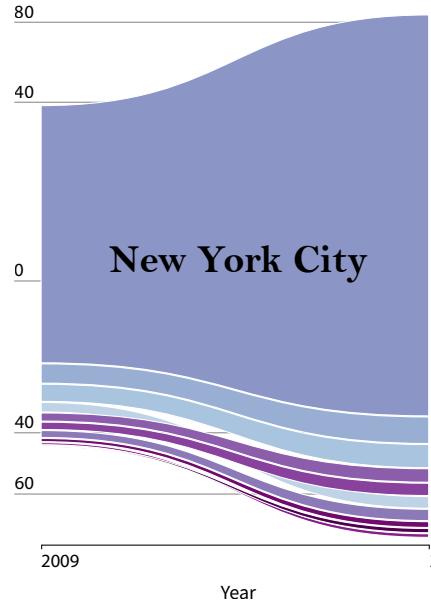
Number of individuals in thousands
The total number of students under homeless conditions and the main differences compared to the base year



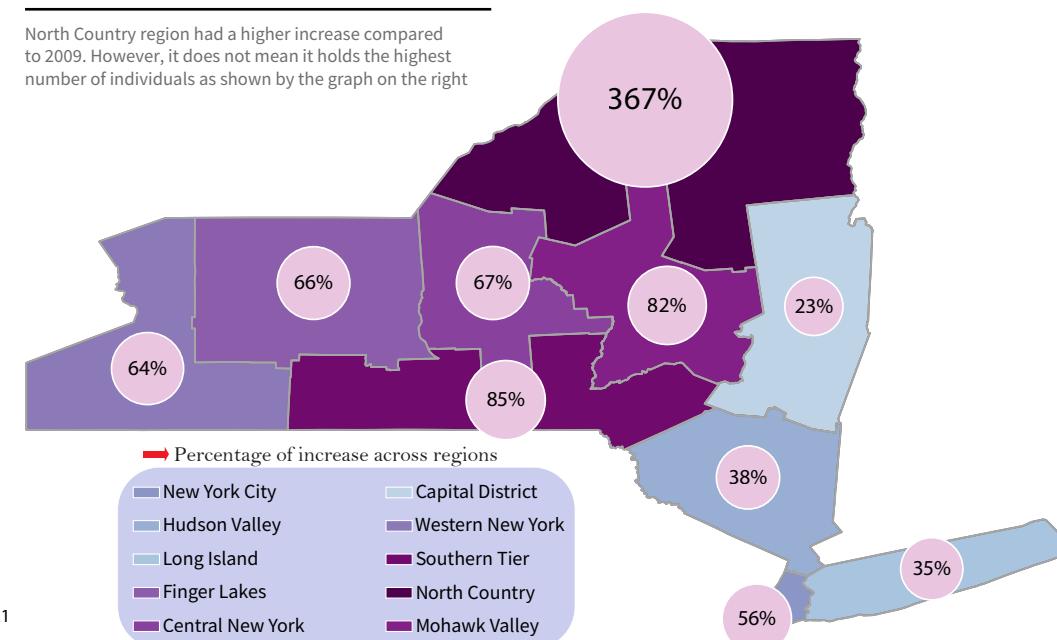
The number of homeless students majorly increased in every county, and just a few have sufficient data



Number of individuals in thousands
Homelessness students increased across all economic regions and were more significant in the New York City region.



North Country region had a higher increase compared to 2009. However, it does not mean it holds the highest number of individuals as shown by the graph on the right



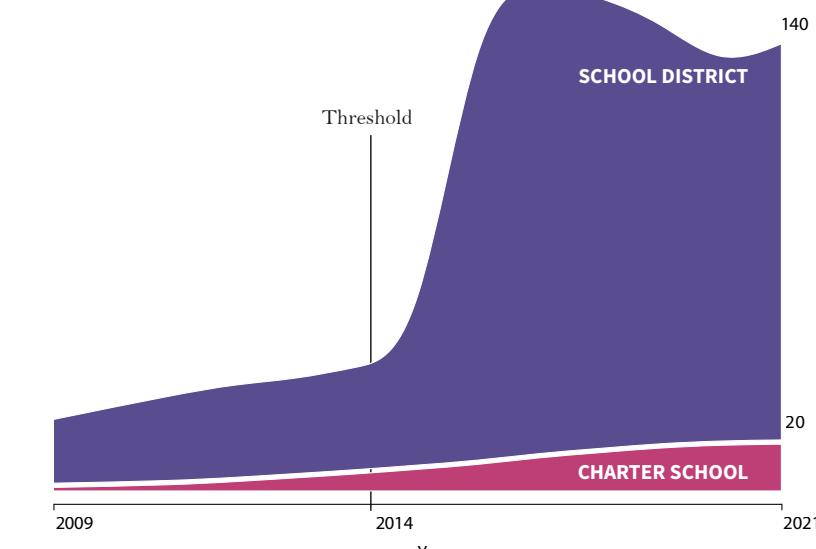
New York State has experienced a considerable increase in students living under homeless conditions over the past 12 years, reaching its apex in 2017. Even though there was a decrease in later years, it was not enough to mitigate the problem and might increase again.

Numbers of increases are in a scale of thousands across all over the state, and here we highlight the proportion of increase from 2009. The raised threshold started in 2014 and mainly impacted public schools, such as Charter Schools and School District, being much higher in the last one. The disproportion was also different

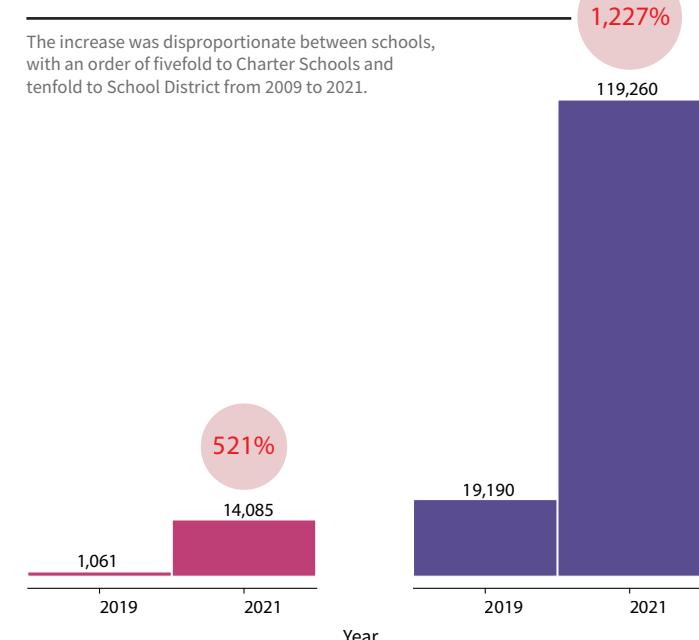
within students' age, as we can see by the number of students per school grade, which reflect higher numbers of homeless for middle and high school students. The *Advocate for Children of New York* reports that 1 in 10 students was homeless in 2021, studying and facing the reality of living under extreme conditions.

Number homeless students in thousands

The more significant part of the homeless students frequents the School District instead of Charter Schools. In both schools, the frequency of these students increases after 2014.

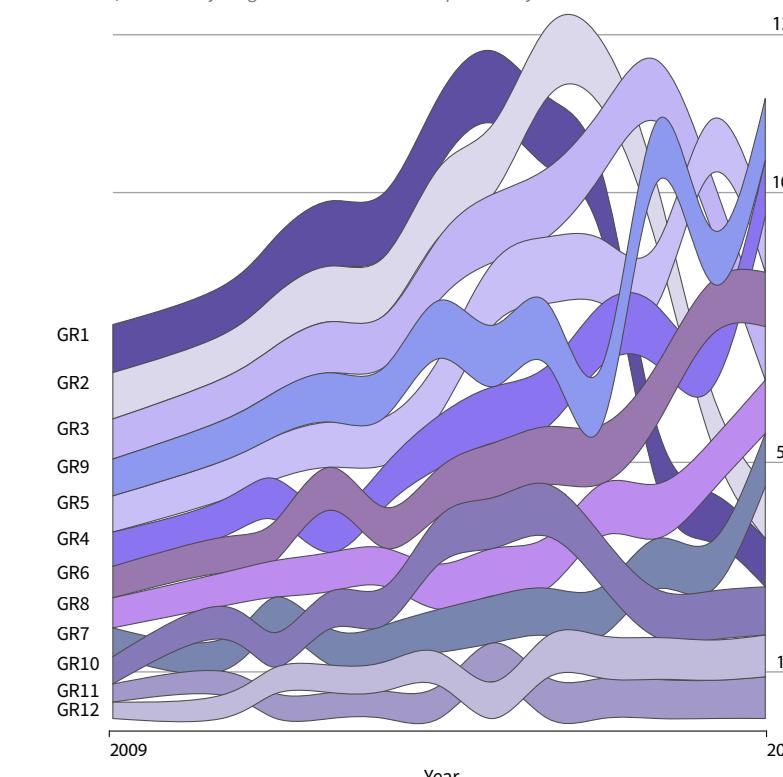


The increase was disproportionate between schools, with an order of fivefold to Charter Schools and tenfold to School District from 2009 to 2021.



Number of homeless students per grade

Over time the increase was constant or slight for older students than younger ones. However, until 2014 younger students increased exponentially



Homeless students' proportion from 2009 to 2021 almost increase at the same rate. Students in primary grades tended to have the same ratio, and the gap between them increased for those attending later grades.

