

JMP Instructional data set 003

Homelessness

JMP platforms and statistical methods

Tabulate:	Table summary statistics
Graph Builder:	Bar chart; line graph; geographic map
Formula Editor:	Creating new variables

Problem statement

It is estimated that over half a million people are homeless on a given night in the United States. Each year, the US Department of Housing and Urban Development (HUD) releases an Annual Homeless Assessment Report to Congress [1]. This report provides estimates of the number and types of people experiencing homelessness across the United States including estimates of how many of those can access a shelter. Producing accurate estimates for this report is a challenging undertaking but is necessary to provide visibility into this problem, identify trends, and ultimately help lead to initiatives that try to address this troubling issue.



The focus of the analyses is to explore trends in the amount of homelessness, examining the rates of those that are sheltered vs. unsheltered, compare homelessness rates for different age groups, and examine/compare these rates at the state level.

Data

IDS_003.jmp

Year	Year of data collection (2013, , 2024)
HUD Region	US Housing and Urban Development Region (I, II, ... , X)
State	State name (including Washington D.C.)
Population	Estimated population of state (including Washington D.C.)
Overall Homeless	Estimate of the total number of people experiencing homelessness on a given night
Overall Homeless_<18	Estimate of the total number of people under the age of 18 experiencing homelessness on a given night
Overall Homeless_18-24	Estimate of the total number of people between the ages of 18 and 24 experiencing homelessness on a given night
Overall Homeless_>24	Estimate of the total number of people over the age of 24 experiencing homelessness on a given night
Sheltered Homeless	Estimate of the total number of people experiencing homelessness on a given night that are staying in a shelter <i>Note: Sheltered means being housed for the night in either an Emergency Shelter, Transitional Housing, or Safe Haven.</i>
Sheltered Homeless_<18	Estimate of the total number of people under the age of 18 experiencing homelessness on a given night that are staying in a shelter
Sheltered Homeless_18-24	Estimate of the total number of people between the ages of 18 and 24 experiencing homelessness on a given night that are staying in a shelter
Sheltered Homeless_>24	Estimate of the total number of people over the age of 24 experiencing homelessness on a given night that are staying in a shelter
Unsheltered Homeless	Estimate of the total number of people experiencing homelessness on a given night that are not sheltered <i>Note: Unsheltered means nighttime location is a public or private place not designed for sleeping accommodations such as on a street, alley, park, vehicle, or abandoned building.</i>
Unsheltered Homeless_<18	Estimate of the total number of people under the age of 18 experiencing homelessness on a given night that are not sheltered
Unsheltered Homeless_18-24	Estimate of the total number of people between the ages of 18 and 24 experiencing homelessness on a given night that are not sheltered
Unsheltered Homeless_>24	Estimate of the total number of people over the age of 24 experiencing homelessness on a given night that are not sheltered

Exercises

1. Create a table showing the count nationally of the number of homeless, sheltered and unsheltered people overall and for each age group.

Instructions:

Analyze > Tabulate. Place the twelve variables 'Overall Homeless', ..., 'Unsheltered Homeless_>24' in the Drop zone for rows. Place 'Year' in the Drop zone for columns.

2. Create a stacked bar chart displaying the count of sheltered and unsheltered homeless for each year. Are there any years where the values are very different than the others? Do you think the numbers for that year are accurate counts? If not, why do you think so? Aside from that year, are there any trends observed? Is so, are they showing positive or negative outcomes?

Instructions:

Graph > Graph Builder. Place the 'Year' variable on the X axis. Place both 'Sheltered Homeless' and 'Unsheltered Homeless' on the Y axis. Select bar chart icon  from the Graph Palette. Choose Stacked Bar Style in the controls on the left. Choose Sum for the Summar Statistic, and Label by Value for the Label Format.

3. Remove the year(s) identified in Exercise 2 from being used in any future analyses.

Instructions:

Select the Filter Views tab in the upper left script panel. Select the "+" symbol. Name the filter 'Years to Exclude'. In the Table Filter panel on the right, choose the variable 'Year' and select the "+" symbol. Choose the year(s) to exclude, then click the Inverse check box. The data table has now been subset with the chosen year(s) excluded. Any analyses performed or graphs created will not include those years. This filter can be turned on/off by selecting it or select "Exit View" in the Filter Views tab.

4. Create similar stacked bar charts as in Exercise 2 – one for each of the three age groups. Comment on any trends observed. Are these trends similar across all three age groups?

5. Create a map of the United States where the color used for the states can be the values for any of the variables in the data set for any year. Exploring specific variables and years, describe 3-5 features in the data the graph reveals.

Instructions:

Graph > Graph Builder. Place 'State' in the center of the graph window. Place 'Overall Homeless' in the Color Role. Then choose Redo > Column Switcher under the Red Triangle. Select 'Overall Homeless' for the Initial Column to Switch. Then select that variable along with the other eleven overall, sheltered, and unsheltered variables for the Replacement Columns. The Column Switcher will appear on the left allowing you to select the specific variable to use. The colors are based on the average values across all the years. To display the values for a single year, set up a Local Data Filter. Select Local Data Filter under the Red Triangle. Select 'Year' as the filter variable. The Data Filter will now appear allowing you to select a specific year to use in the graph. You'll notice that the default scale to use for the color variable makes it difficult to compare states as states with very large population sizes (e.g., California, New York) will have large values for all these variables resulting in the remaining states all being similar color. To remedy this, right click on the color scale and choose Gradient. Select Quantile as the Scale Type. Now the median (50th percentile) is grey, and the colors are based upon their percentile not actual values.

6. Create a new variable that is the Overall Homeless per 100K people.

Instructions:

Create a new column (Cols > New Columns). In the New Column dialog box, choose Formula in the Column Properties drop down menu. Select Edit Formula. In the Formula Editor, create the formula:

$$\left(\frac{\text{Overall Homeless}}{\text{Population}} \right) \cdot 100000$$

Name the column 'Overall Homeless per 100K'.

7. Create a map of the United States where the color used for the states is the overall homeless per 100K people, with the ability to select a given year to use for the display. Examine how the values change for the states across the years. Describe 2-3 features in the data the graph reveals.

Instructions:

Graph > Graph Builder. Place 'State' in the center of the graph window. Use 'Overall Homeless per 100K' as the Color Role. Select Local Data Filter under the Red Triangle. Select 'Year' as the filter variable. Right click on the color scale and choose Gradient, and select Quantile as the Scale Type.

8. Create an overlay line graph displaying the 'Overall Homeless per 100K' for each State across the years, with the ability to select specific HUD Regions and/or States to display. Select one of the HUD Regions. Compare the States in that region and how the homelessness rates change over time. Describe 2-3 features about that region the graph reveals.

Instructions:

Graph > Graph Builder. Place 'Year' in the X Role. Place 'Overall Homeless per 100K' in the Y Role. Select Line Graph from the graph palette. 

Place 'State' in the Overlay role. Select Local Data Filter under the Red Triangle. Select 'HUD Region' and 'State' as the filter variables. The Data Filter will now appear allowing you to select specific HUD Region(s) and/or State(s) to display the graph.

9. What are some other data that would be valuable to analyze to help develop a deeper understanding of this issue possibly helping identifying potential causes?

Supporting materials

1. The 2024 Annual Homelessness Assessment Report (AHAR) to Congress.
Part 1: Point-In-Time Estimates of Homelessness. Dec 2024.
<https://www.huduser.gov/portal/sites/default/files/pdf/2024-AHAR-Part-1.pdf>

Data Source: <https://www.huduser.gov/portal/datasets/ahar/2024-ahar-part-1-pit-estimates-of-homelessness-in-the-us.html>

** Please adhere to any citation requirements for distribution and use of these data.*