

Homework 8

Due: Monday April 15 at 11:59pm in compass2g.illinois.edu

Use RStudio for all exercises. Provide one code file that contains all the code and includes comments noting which code is for which exercises. You will also need to comment on the results, so place them in a Word (or Open Office or HTML or PDF) document and add your comments to answer the questions. You can use knitr to programmatically create your document or use an R notebook if you like. Be sure to include your name in the file name for all files submitted.

Any code based on code from elsewhere must reference in comments the source of the original code. **Code files must be the actual code files**, not code pasted into some other document.

All exercises are based on the **USArrests** dataset from R. This data set contains some state-wise arrest statistics as well as urban populations for all states in the US in 1973. Recall that some plotting functions expect a matrix and some expect a data frame, so you will need to use modified copies of the data when necessary.

Use options to add clear labels and titles to your plots, and avoid the default `data$variable` labeling in your plots.

Exercise 1:

Obtain a heatmap based on the **USArrests** data set, and interpret the plot.

Specifically, comment on which states have higher and lower values of these statistics, comment on any apparent relationships between these crime rates and urban population, and comment on groups of states that are the most similar with respect to these statistics and groups of states that are very different with respect to this statistics.

Exercise 2:

The **state** data set in R contains some basic information about the states in the United States in 1975. The states are in alphabetical order in the **state** data set and in the **USArrests** data set, so it is easy to directly combine information from them.

Combine the **USArrests** data with the region and abbreviation (`abb`) variables from the **state** data set to create a data set for the treemap exercises.

Use a treemap to compare state-by-state urban population and assault rates, grouping by region within the plot.

Comment on urban populations and assault rates by region and within region. What were the general differences between regions at the time? What were the differences within regions? Which states had the highest and lowest urban populations and assault rates overall in 1973, and which were highest and lowest within regions.

Exercise 3:

Repeat Exercise 2 with murder and assault rates instead of urban population and assault rates.

Exercise 4:

Create visualizations of the following three kernel density estimations.

- Murder rate and assault rate
- Murder rate and urban population
- Assault rate and urban population

Be sure to choose bandwidths that provide reasonable refinement for the density estimation (not overly smoothed or under-smoothed). Comment on what the estimations tell us about more common magnitudes of the arrest rates and urban populations at the time, and what the plots tell us about relationships between the arrest rates and between the arrest rates and urban population in 1973.

Additional Exercise for Graduate Students

Exercise 5:

Obtain and compare visualizations of murder and assault rate kernel density estimations by region using the region variable added for the treemaps. Comment on how the arrest rates differ across regions and how the individual regions compare with the overall density estimation for murder and assault rates from Exercise 4.

Get the plots by group programmatically rather than doing each plot in separate replicated code.