Assignment 7: US Counties Data Visualization

30 Points scaled to 20 Points

Introduction

This assignment asks you to generate a series of graphs relating to county-level data for the contiguous United States. You will make use of the "us_counties_data.csv" dataset. Make sure to read the associated description of the data ("us_counties_data_ DESCRIPTION.pdf") prior to undertaking the assignment. The description file explains the fields available and also provides the data source. This exercise will also require you to subset and query the data as not all graphs will use all of the data points.

Objectives

- *Query, filter, and summarize data*
- *Use a variety of graphical parameters to visualize and compare variables*
- *Generate effective and informative graphics*

Deliverables

- *Jupyter Notebook (Python) or R Markdown file (R) with all code and graphs embedded. Files can be rendered to HTML webpages if your instructor requires this. Graph prompts should be stated within Markdown cells.*

Graphs

This assignment can be conducted using either Python (matplotlib, seaborn, and pandas) or R (ggplot2), whichever you prefer or whichever you instructor requires. Generate code to create the following graphs.

Graph 1: Generate a box plot showing the distribution of county-level median income ("med_income") for all counties in the dataset. (2 Points)

Graph 2: Generate a grouped box plot showing the distribution of county-level median income ("med_income") grouped by subregion of the country ("SUB_REGION") for all counties in the dataset. (2 Points)

Graph 3: Generate a histogram showing the distribution of county-level median income ("med_income") for all counties in the dataset. (2 Points)

Graph 4: Generate a kernel density plot showing the distribution of county-level median income ("med_income") for all counties in the dataset. (2 Points)

Graph 5: Generate a histogram showing the distribution of county-level percent forest cover ("per_for") for only counties occurring in the West North Central ("W N Cen") subregion ("SUB_REGION"). (2 Points)

Graph 6: Generate grouped kernel density plots showing the distribution of county-level percent cropland ("per_crop") for only counties occurring in the West North Central ("W N Cen"), Pacific ("Pacific"), and Mountain ("Mtn") subregions ("SUB_REGION"). Each subregion should have a separate density plot in the graph space. (2 Points)

Graph 7: Generate a dot plot showing the mean percent crop land ("per_crop") for only counties occurring in the West North Central ("W N Cen"), Pacific ("Pacific"), and Mountain ("Mtn") subregions ("SUB_REGION"). Three data points should be provided: one for each subregion. (2 Points)

Graph 8: Add error bars to Graph 7 showing 1 standard deviation above and below the mean for each subregion separately. (2 Points)

Graph 9: Create a scatter plot showing the relationship between mean elevation ("dem") and mean county-level average annual temperature ("tempmn") for all counties in the country. (2 Points)

Graph 10: Replicate Graph 9 but now differentiate the subregions ("SUB_REGION") by using different colors or point symbols. In other words, all counties in the same subregion should have the same color or symbol applied. (2 Points)

Graph 11: Create a scatter plot showing the relationship between mean elevation ("dem") and mean county-level average annual temperature ("tempmn") for only counties occurring in the West North Central ("W N Cen"), Pacific ("Pacific"), and Mountain ("Mtn") subregions. Differentiate the subregions ("SUB_REGION") by using different colors or point symbols. In other words, all counties in the same subregion should have the same color or symbol applied. (2 Points)

Graph 12: Generate a grouped box plot showing the distribution of mean county elevation ("dem") grouped by state ("STATE_NAME" or "STATE_ABBR") for only counties occurring in Florida, Colorado, Montana, and Nebraska. (2 Points)

Graph 13: Generate a scatterplot for only counties occurring in Vermont to compare the population density ("POP_SQMI") and mean county-level average annual temperature ("tempmn"). Label each point in the plot with the county name. (2 Points)

Graph 14: Generate a scatterplot showing the relationship between the density of railroads ("rail_den") and roads ("road_den") for just counties occurring in the New England ("N Eng") subregion ("SUB_REGION"). Also, use the size of the point symbol to display percent developed land ("per_dev"). (2 Points)

Graph 15: Generate a scatterplot showing the relationship between the density of railroads ("rail_den") and roads ("road_den") for just counties occurring in the New England ("N Eng") subregion ("SUB_REGION"). Also, use the point color or point symbol to differentiate counties by state. (2 Points)