Example Assignment

I chose to analyze a dataset on dog heartworm cases, obtained from the Companion Animal Parasite Council website: <https://capcvet.org>. Specifically, I obtained annual, county-level counts of dog heartworm cases in California from 2012 – 2021. I was interested in whether there were trends in dog heartworm case counts over time, and between different biogeographic regions of California. I used the California bioregion designations developed by the Interagency Natural Areas Coordinating Committee (Figure 1a), then used the aggregate function to sum the case counts for each year and bioregion (Figure 1b). For Figure 1a, I used the *tigris* package, which downloads shapefiles for U.S. Census regions. I used the *RColorBrewer* package to generate the colors used for each bioregion. For Figure 1b, I used ggplot.

**Figure 1.**

**Chart, line chart

Description automatically generated**

I was then interested in whether there was any relationship between dog heartworm cases and socioeconomic status, because I know that heartworm preventatives are effective but can be prohibitively expensive. For this analysis, I downloaded data on median household income from the U.S. Census Bureau: <https://www.census.gov/topics/income-poverty/income.html>. Using this data, I ran a linear regression in R using dog heartworm cases in a given year and county as the outcome variable and median household income as the predictor. I found there was a significant, positive relationship between case counts and income (Figure 2).

**Figure 2.**

.Chart, scatter chart

Description automatically generated