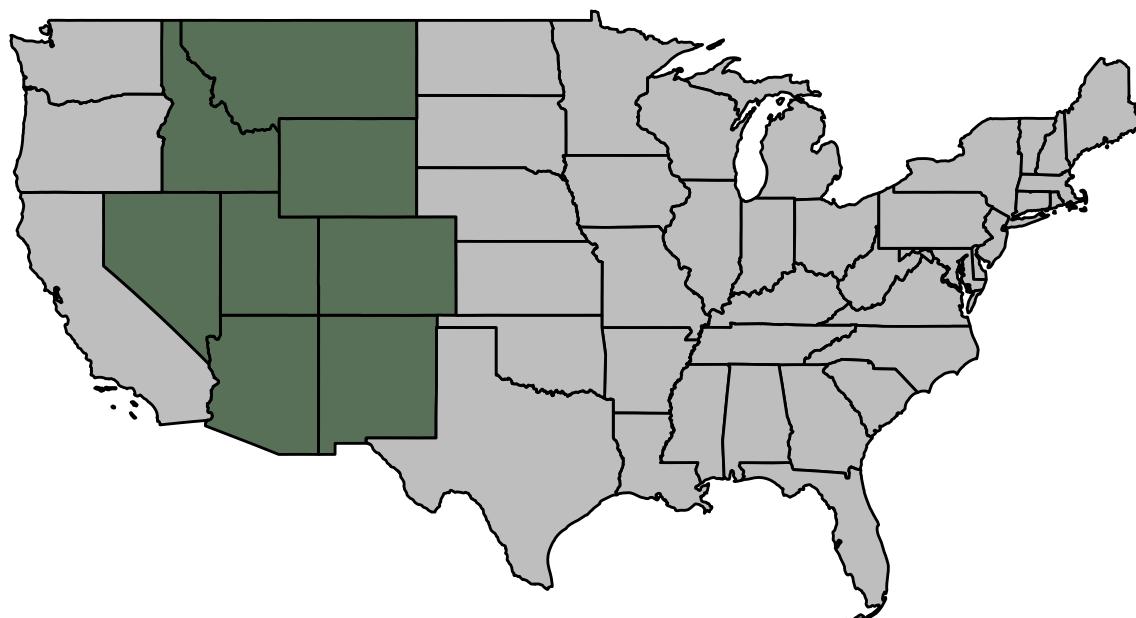


## Data

The data used in this thesis was collected by the Forest Inventory and Analysis Program (FIA) in the span of 10 years from 2007 to 2017. While this data was collected over this 10 year period, the analyses done throughout this thesis are under the assumption that this is a “snapshot” of the Interior West at some moment in time. The data we have is plot-level data for the Interior West region of the United States, where the data for each plot is collected by FIA and its crew members. The units measured by the FIA and their ground crews are approximately 30 m by 30 m hexagonal units. Since the Interior West covers over 140 million acres of forestland, it is extremely impractical for FIA to measure every unit (SOURCE: <https://www.fs.fed.us/rm/ogden/lib/interiorwest2.pdf>). Instead, they sample from the population of 30 m by 30 m hexagonal units by using a geographically-based systematic sampling design. The FIA chooses these samples by first overlaying a hexagonal grid over the United States where each hexagon contains 6000 acres of land. Then, they fill these hexagons with much smaller hexagons and randomly sample from the population of small hexagons. Then, ground crews go to these sampled small hexagons and collect variables such as basal area, trees per acre, etc.

Again, the data we have is from the Interior West, and the FIA defines the Interior West as Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming. For reference we have provided the Interior West colored green on a map of the continental United States:



While the data covers the Interior West as a whole, we have very granular information, as each row represents a plot sampled by FIA. The data also includes variables that subset the Interior West into provinces which contain ecoregions, and these ecoregions contain ecoregions. In our data, on average, each ecoregion contains approximately 7.06 ecoregions, and each province contains an average of 4.86 ecoregions. So, an average province then contains just over 34 ecoregions. This hierarchical structure of the data allows for us to more easily create hierarchical models which borrow strength from surrounding areas.

While this data contains a multitude of variables, the analyses done in this thesis focus on four key response variables and two explanatory variables. The response variables used are basal area (square-foot), trees per acre, above-ground biomass (lbs), and net volume ( $\text{ft}^3$ ). The explanatory variables used are remotely sense variables: forest biomass and forest probability. We can look at the average of these variables across the Interior West region by ecosubsection:

