<https://archive.ics.uci.edu/ml/datasets/Covertype>

Regarding covtype.info (the readme for the dataset):

This dataset contains 581,012 instances with 12 measures across 54 columns of data. 40 of these columns are binary (boolean) variables for soil type, and 4 more columns are binary variables indicating which wilderness area. That leaves us with 10 core quantitative variables, but the 44 columns of binary could be converted to a 2 bit variable for the wilderness area indicator and 6 bits for the soil types. I may choose to do this, but perhaps working with the one-hot encoding for soil type will help preserve uncertainty. Soil type is a categorical variable and not quantitative, so putting them on a numbered scale may not necessarily make sense.

Elevation quantitative meters Elevation in meters

Aspect quantitative azimuth Aspect in degrees azimuth

Slope quantitative degrees Slope in degrees

Horizontal\_Distance\_To\_Hydrology quantitative meters Horz Dist to nearest surface water features

Vertical\_Distance\_To\_Hydrology quantitative meters Vert Dist to nearest surface water features

Horizontal\_Distance\_To\_Roadways quantitative meters Horz Dist to nearest roadway

Hillshade\_9am quantitative 0 to 255 index Hillshade index at 9am, summer solstice

Hillshade\_Noon quantitative 0 to 255 index Hillshade index at noon, summer soltice

Hillshade\_3pm quantitative 0 to 255 index Hillshade index at 3pm, summer solstice

Horizontal\_Distance\_To\_Fire\_Points quantitative meters Horz Dist to nearest wildfire ignition points

Wilderness\_Area (4 binary columns) qualitative 0 (absence) or 1 (presence) Wilderness area designation

Soil\_Type (40 binary columns) qualitative 0 (absence) or 1 (presence) Soil Type designation

Cover\_Type (7 types) integer 1 to 7 Forest Cover Type designation