

CPE 232 Data Models

HW 4: Data Visualization and EDA

Use the `covertype` dataset and plot an appropriate visualization to answer the following questions:

- Distribution of `cover_type` across different `wilderness_type`
- Distribution of elevation on each cover type
- Number of data points on each climatic zone
- Which cover type dominates subalpine climate?
- Which cover type is likely to be found near a body of water?
- Which cover type is likely to be found in a shady area from 9 am to noon?
- Which cover type is likely to be found away from hazardous points (roadways/firepoints)?
- Determine whether elevation provides more information to distinguish `Krummholz` from `Cottonwood/Willow` than `horizontal_distance_to_hydrology`?
- Compare the relationship between `Krummholz` and `Cottonwood/Willow` to geographical features (elevation, distance to hydrology, hillshade, etc.).
- Plot histogram of `elevation`, `aspect`, and `slope`. Discuss the shape of the distribution.
- For each member in your group, plot another useful visualization from this dataset.