

```

{
  "clusters": [
    {
      "name": "NIFC Wildfire Statistics",
      "files": [
        "nifc_wildfires.csv",
        "nifc_lightning_caused_wildfires.csv",
        "nifc_human_caused_wildfires.csv",
        "nifc_lightning_caused_acres.csv",
        "nifc_human_caused_acres.csv",
        "nifc_suppression_costs.csv",
        "nifc_geographic_areas.gpkg"
      ],
      "description": "Comprehensive wildfire data from the National Interagency Fire Center (NIFC).",
      "reason": "All files are sourced from NIFC (as indicated by the 'nifc_' prefix) and provide detailed statistics on wildfires, including their causes (human vs. lightning), impact (acres burned), suppression costs, and geographic boundaries."
    },
    {
      "name": "NOAA Wildfire Data",
      "files": [
        "noaa_wildfires.csv",
        "noaa_wildfires_monthly_stats.csv",
        "noaa_wildfires_variabledescrip.csv"
      ],
      "description": "Wildfire data and metadata from the National Oceanic and Atmospheric Administration (NOAA).",
      "reason": "These files form a cohesive dataset from NOAA (as indicated by the 'noaa_' prefix), including the main data, monthly statistics, and a description of the variables used."
    },
    {
      "name": "State-Level Wildfire Statistics",
      "files": [
        "Wildfire_Acres_by_State.csv",
        "Wildfire_Fires_by_State.csv",
        "wildfire_total_fires_p45_54.csv"
      ],
      "description": "Aggregated statistics on the number of wildfires and acres burned, primarily categorized by state.",
      "reason": "These files all provide high-level, aggregated wildfire statistics, with a focus on state-by-state breakdowns as suggested by their names."
    },
    {
      "name": "Wildfire Environmental and Operational Data",
      "files": [
        "Fire_Weather_Data_2002-2014_2016.csv",
        "PublicView_RAWS_-3515561676727363726.csv",
        "annual_aqi_by_county_2024.csv",
        "cleaned_helicopter_requests_by_region.csv"
      ],
      "description": "Data related to the environmental conditions for wildfires and the operational response.",
      "reason": "This cluster groups files about factors influencing wildfires (weather data from RAWs), their consequences (air quality), and the operational response (helicopter usage)."
    },
    {
      "name": "US General Geographic and Demographic Data",
      "files": [
        "usa.gpkg",
        "NST-EST2024-POP.xlsx",
        "democratic_vs_republican_votes_by_usa_state_2020.csv",
        "state_abbreviation_to_state.json"
      ],
      "description": "General reference data for the United States, including geographic boundaries, population estimates, political data, and state name mappings.",
      "reason": "These files provide contextual geographic, demographic, and political information for the USA, which can be used to enrich the primary wildfire datasets."
    },
    {
      "name": "Utility Scripts",
      "files": [
        "load_data.py"
      ],
      "description": "A code script for data processing.",
      "reason": "This is a standalone Python script, likely used to load or manipulate the other data files in the project."
    }
  ]
}

```