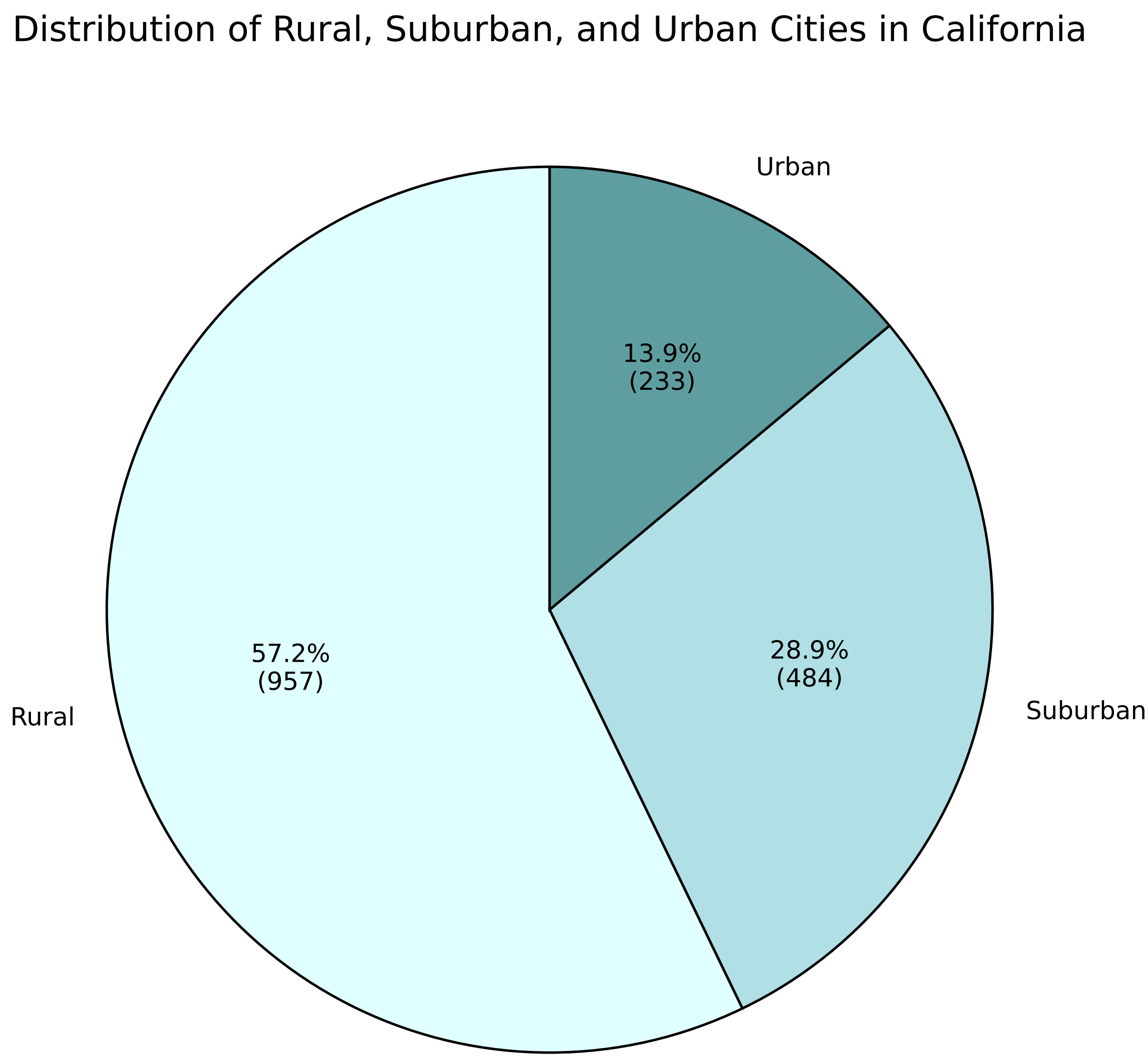
# Final Project: Writeup

Jae Hu (jianinghu0408), Duoshu Xu (KevinX0), Regina Hou (Reginahk), Section 3

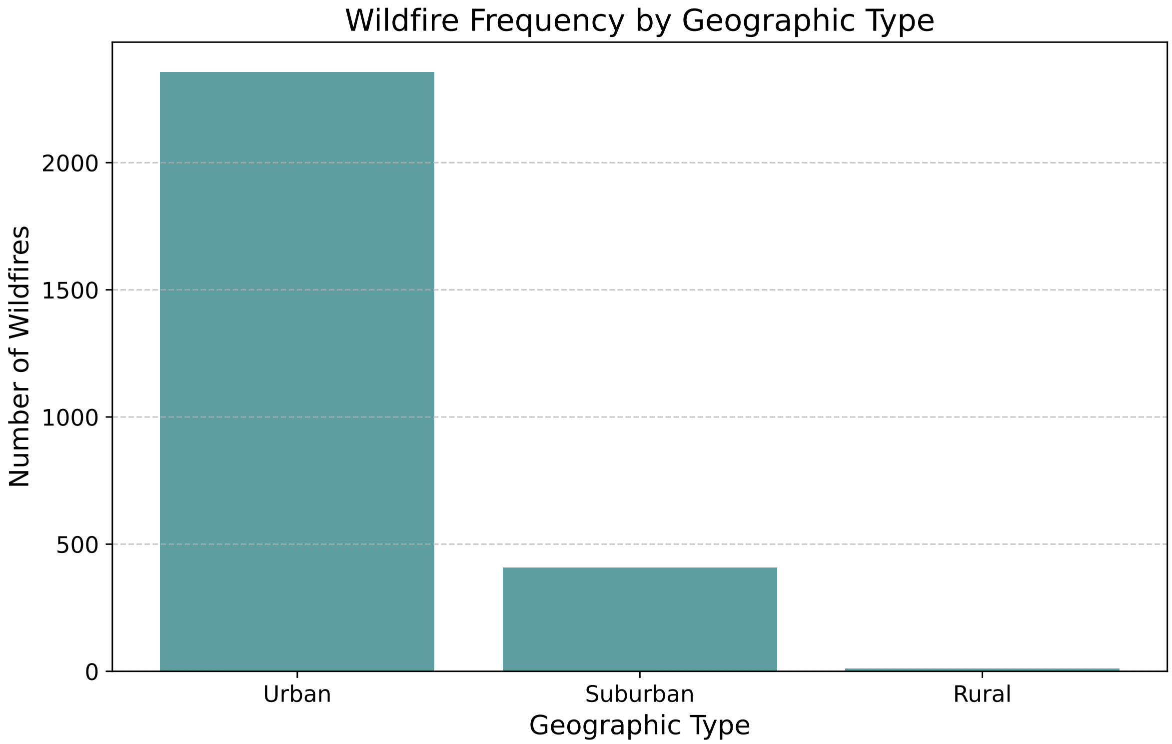
## Data cleaning and reshaping

## pie chart (done by Jae Hu)



## Wildfire Frequency bar chart (done by Duoshu Xu)

```
  Geographic Type  Number of Wildfires
0           Urban                 2356
1        Suburban                  407
2           Rural                   10
```




```
['California' 'Alameda County' 'Alameda' 'Albany' 'Ashland CDP' 'Berkeley'
 'Castro Valley CDP' 'Cherryland CDP' 'Dublin' 'Emeryville']
```

```
/var/folders/r4/x5b99tvj66zcn_88m3jn4r6w0000gn/T/ipykernel_23835/3288246651.py:4:
DtypeWarning:

Columns (12,36) have mixed types. Specify dtype option on import or set
low_memory=False.
```

| | Geography | Total Population | Land Area in Square Miles | Population Per Square Mile (Land Are |
|---|---|---|---|---|
| 0 | California | 39538223 | 155858.326771 | 253.680530 |
| 1 | Alameda County | 1682353 | 737.461854 | 2281.274605 |
| 2 | Alameda | 78280 | 10.448679 | 7491.856487 |

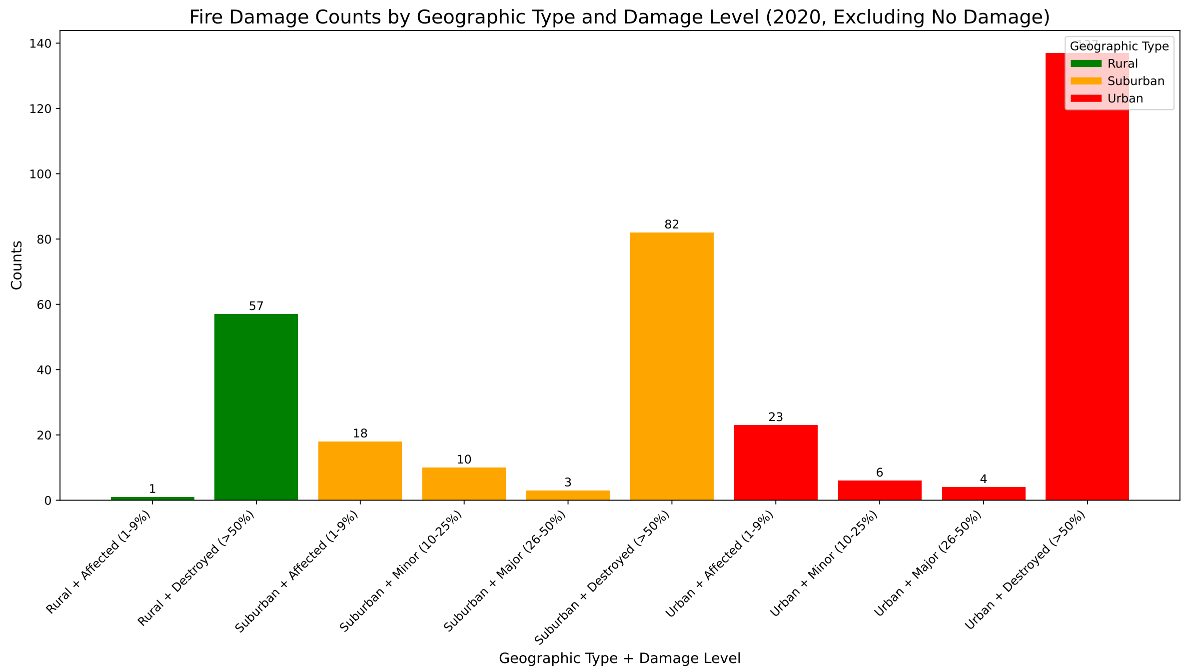
| | Geography | Total Population | Land Area in Square Miles | Population Per Square Mile (Land Are |
|---|---|---|---|---|
| 3 | Albany | 20271 | 1.789982 | 11324.694641 |
| 4 | Ashland CDP | 23823 | 1.842571 | 12929.213531 |

| | OBJECTID | * Damage | * Street Number | * Street Name | * Street Type (e.g. road, drive, lane, e |
|---|---|---|---|---|---|
| 0 | 1 | No Damage | 8376.0 | Quail Canyon | Road |
| 1 | 2 | Affected (1-9%) | 8402.0 | Quail Canyon | Road |
| 2 | 3 | No Damage | 8430.0 | Quail Canyon | Road |
| 3 | 4 | No Damage | 3838.0 | Putah Creek | Road |
| 4 | 5 | No Damage | 3830.0 | Putah Creek | Road |


| | OBJECTID | * Damage | * Street Number | * Street Name | * Street Type (e.g. road, drive, lane, e |
|---|---|---|---|---|---|
| 0 | 1 | No Damage | 8376.0 | Quail Canyon | Road |
| 1 | 2 | Affected (1-9%) | 8402.0 | Quail Canyon | Road |
| 2 | 3 | No Damage | 8430.0 | Quail Canyon | Road |
| 3 | 4 | No Damage | 3838.0 | Putah Creek | Road |
| 4 | 5 | No Damage | 3830.0 | Putah Creek | Road |

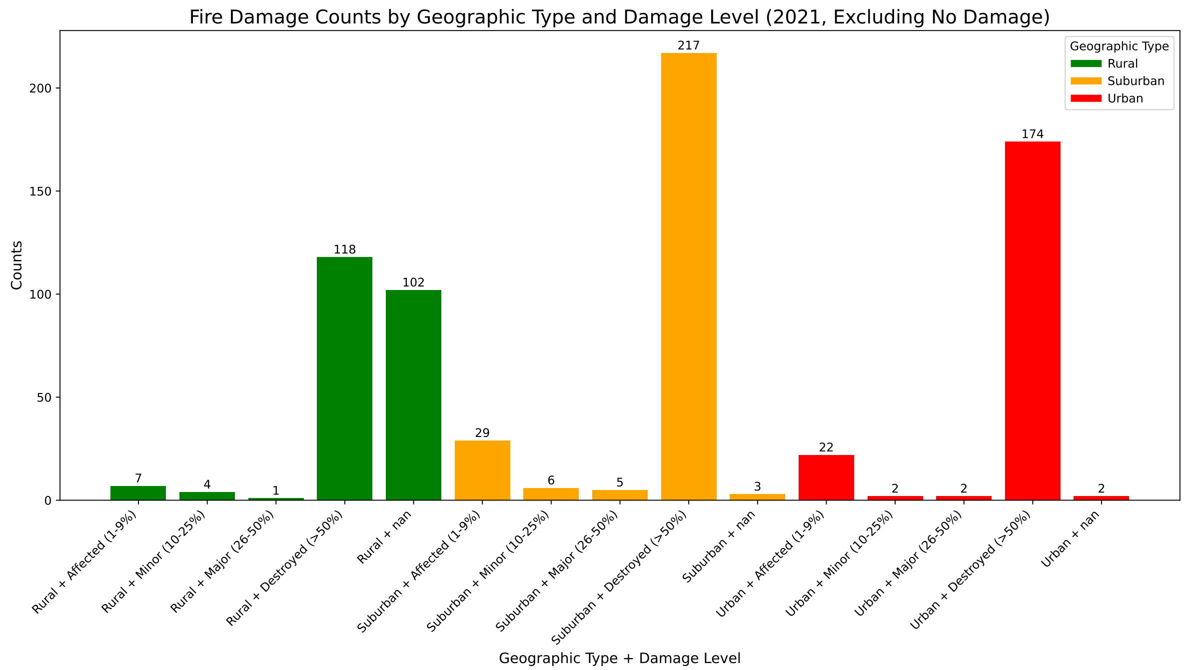

## bar plot of 2020 (done by Regina Hou)

```
/var/folders/r4/x5b99tvj66zcn_88m3jn4r6w0000gn/T/ipykernel_23835/454140121.py:2:
UserWarning:

Could not infer format, so each element will be parsed individually, falling
back to `dateutil`. To ensure parsing is consistent and as-expected, please
specify a format.

/var/folders/r4/x5b99tvj66zcn_88m3jn4r6w0000gn/T/ipykernel_23835/3001088990.py:3:
UserWarning:

Could not infer format, so each element will be parsed individually, falling
back to `dateutil`. To ensure parsing is consistent and as-expected, please
specify a format.
```



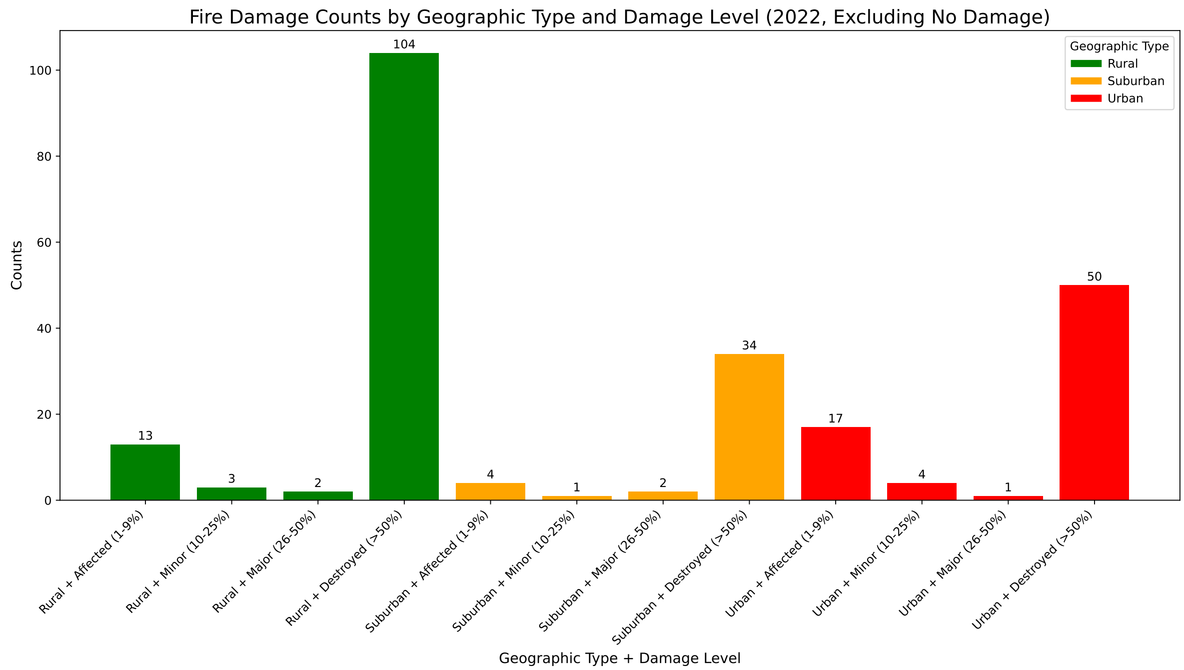

## plot of 2021 (done by Regina Hou)

```
NaN values in geographic type: 0
NaN values in * Damage: 0
```




## plot of 2022 (done by Regina Hou)

```
/var/folders/r4/x5b99tvj66zcn_88m3jn4r6w0000gn/T/ipykernel_23835/3282159924.py:2:
UserWarning:

Could not infer format, so each element will be parsed individually, falling
back to `dateutil`. To ensure parsing is consistent and as-expected, please
specify a format.
```

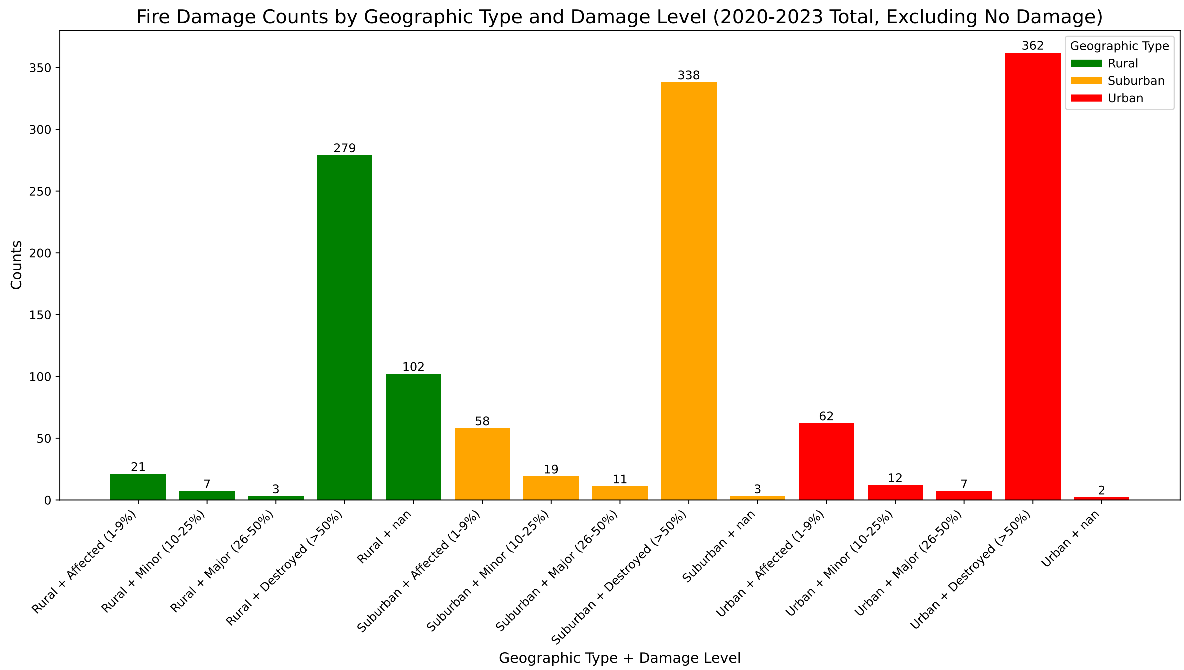



## plot of 2023 (done by Regina Hou)

```
/var/folders/r4/x5b99tvj66zcn_88m3jn4r6w0000gn/T/ipykernel_23835/897568859.py:2:
UserWarning:

Could not infer format, so each element will be parsed individually, falling
back to `dateutil`. To ensure parsing is consistent and as-expected, please
specify a format.
```

## plot for 2020-2023 (done by Regina Hou)

```
/var/folders/r4/x5b99tvj66zcn_88m3jn4r6w0000gn/T/ipykernel_23835/857330891.py:1:
UserWarning:

Could not infer format, so each element will be parsed individually, falling
back to `dateutil`. To ensure parsing is consistent and as-expected, please
specify a format.
```




Intro Wildfires are among the most destructive natural disasters in the U.S., with California experiencing the greatest impact. In 2021, the state accounted for 40% of all burned acres nationwide and had nearly two million properties at risk of wildfire damage—three times more than the next highest state. The rising frequency and severity of wildfires, driven by climate change, droughts, and urban expansion into fire-prone areas, highlight the urgent need to address wildfire risks.

Research question, approach and coding Our study tries to answer the question: "How does the degree of urbanization in areas affected by wildfires relate to the magnitude of property damage? This paper uses data from 2020 to 2023, which allows for analysis of seasonal changes in wildfire incidents. Using cities as a unit of analysis, we divided the cities into urban, suburban, and rural areas based on the population density. The findings of this study offer valuable guidance for tailoring wildfire prevention and response strategies to the unique needs of different regions. To answer this research question, we first cleaned and integrated our three datasets. We standardized geographic names by removing redundant text, added a"Geographic Type" column to classify cities based on their population, and merged datasets to link wildfire incidents with urbanization levels. Using Shiny App, we created dynamic visualizations and heatmaps to highlight trends and spatial patterns. We encountered a few challenges during the process. One was ensuring that the geographic classifications were accurate, because relying only on population may oversimplify the complexities of urbanization. Second, merging datasets required significant effort to align different formats and resolve inconsistencies in naming conventions. Despite these challenges, our approach provided a robust framework for analyzing wildfire impacts.

Data and methodology We use three data sets in this research. The first one is the 2020 U.S. Census, which provides detailed population data at the city level. It allows for the classification of cities into urban, suburban, and rural categories based on population density. The second dataset is the incident data from the Cal Fire report (mapdataall.csv), which contains incident data related to wildfires. Finally, we use the Cal fire inspection data. This dataset contains the fire damage level with geographic information for us to analyze.

Static plots 1. Pie Chart: Geographic Breakdown of California Cities This chart illustrates the distribution of California cities by geographic type. Rural areas dominate, making up 57.2% of cities, followed by suburban areas at 28.9%, and urban areas, which represent only 13.9% of the total. 2. Bar Chart: Wildfire Frequency by Geographic Type This chart shows wildfire frequency in urban, suburban, and rural areas. Urban areas experience significantly more wildfires than suburban or rural regions 3. Bar Chart: Fire Damage Counts by Geographic Type and Damage Level (2020) This plot breaks down wildfire property damage in 2020 across geographic types and damage levels. Urban areas exhibit the highest counts in severe damage categories, while rural areas see minimal damage. 4. Bar Chart: Fire Damage Counts by Geographic Type and Damage Level (2021) Similar to the previous chart, this one focuses on 2021 and shows a similar pattern: urban areas experience the most significant damage levels, while suburban and rural areas see lower but noticeable impacts. 5. Bar Chart: Fire Damage Counts by Geographic Type and Damage Level (2022) Similar to the previous chart, this one focuses on 2022 and shows a similar pattern 6. Bar: Fire Damage Counts by Geographic Type and Damage Level (2020-2023 Total) Aggregating data from 2020 to 2023, this chart provides a comprehensive overview of damage counts. Urban areas again dominate the severe damage categories, with suburban areas following and rural regions experiencing the least damage overall.

Shiny App Our Shiny App presents a dynamic heatmap of California wildfire damage categorized by geographic type. Users can filter by damage levels and urbanization types to explore patterns interactively. For this study, we focused on the “destroyed” damage level, represented by red dots. The map reveals that “destroyed” damage in suburban areas is more geographically dispersed, while in urban areas, it is concentrated in specific high-risk locations. This visualization helps illustrate the distinct impact of wildfires across different regions.

Policy implications To address wildfire risks effectively, we recommend that for urban areas, the government should focus resources on high-risk zones with severe damage, implementing advanced fire suppression systems and strict building codes to prevent catastrophic losses. Community awareness campaigns should educate urban residents on fire prevention, safe evacuation, and early reporting to reduce wildfire ignition and spread. For suburban areas, land use planning must prioritize vegetation management around suburban developments and regulate construction in vulnerable areas. The government should also establish local volunteer firefighting teams equipped with wildfire suppression tools to enhance community resilience. For rural areas, the government should promote fire-resistant agriculture, such as controlled grazing or fire-resistant crops to create natural barriers to mitigate fire spread.

Limitations This research provides valuable insights into wildfire property damage across different levels of urbanization but has several limitations. First, the classification of geographic areas into rural, suburban, and urban categories based solely on population density may oversimplify the complexity of urbanization. Factors such as infrastructure, land use, and economic activity, which also influence wildfire vulnerability and damage, are not accounted for in this analysis. Second, the broad fire damage categories (e.g., "Minor" for 10–25% damage) limit the precision of analysis and resource allocation. Third, the 2020–2023 timeframe may not capture long-term wildfire trends, therefore limiting the findings' generalizability. Finally, focusing solely on property damage neglects broader wildfire impacts, such as environmental, public health, and social consequences. Future research should address these gaps for more comprehensive insights.

Directions for future work Future work should aim to address the limitations of this study by incorporating more nuanced measures of urbanization, including infrastructure and land use, to capture a comprehensive picture of wildfire vulnerability. Expanding the analysis to include longer time frames would help identify broader trends and improve the generalizability of findings. Additionally, integrating data on the social, economic, and ecological impacts of wildfires—such as public health effects, environmental degradation, and community displacement—would provide a more holistic understanding of wildfire consequences and inform more effective, multidimensional policy strategies.

Reference: Heacock, D. (2022). U.S. states most impacted by wildfires. FilterBuy.com. https://filterbuy.com/resources/across-the-nation/states-impacted-by-wildfires/ NASA. (2021, October 5). What's behind California's surge of large fires? NASA. https://earthobservatory.nasa.gov/images/148908/whats-behind-californias-surge-of-large-fires