Primary Cause Of Big Wildfires In Each US State And Each US Region

By: Tilak Muley

Initial problem/question statement

What is the primary cause of big wildfires in each US state, and when the states are grouped by region which causes are the most primary?

Abstract

Wildfires pose a great threat to the environment and humans, so understanding the causes and areas prone to these causes of wildfires is greatly needed to help not only prevent wildfire in the future but to also manage these areas. This project's goal is to investigate the primary cause of the biggest wildfires in each US state from 1992-2020 and to see any correlation between regions and primary causes. The causes are categorized typically by natural or human but for this project it is going to go into more categories like debris and open burning, arson/incendiarism, equipment and vehicle use, recreation and ceremony, misuse of fire by a minor, smoking, railroad operations and maintenance, fireworks, power generation/transmission/distribution, firearms and explosives use and nature. This project will research into how different regions have different causes to the biggest wildfire experienced by them, so regions can be better prepared to fight wildfires based on the cause of the biggest wildfire. As there is a reason most of the time why a big wildfire based on size was able to burn so much typically the cause of it has something to do with it. Knowing the primary causes of wildfires of a region can help make plans to not only prevent but also control and take down wildfires faster. Which in turn can help save infrastructure and lives. The findings in this project can have implications on environmental policies, how the land is used and to have the public more aware on wildfires and the cause in the region they call home. It is found that different regions have different wildfire sizes and that causes of big wildfires in each region does add up according to research done and the findings in this project.

Introduction

Wildfires are a big concern in the United States as they cause a lot of damage to ecosystems, the land, and properties and not to forgot lives can be lost in wildfires. Wildfires are uncontrolled, unplanned and unpredictable fires. Primary causes of wildfires that are assumed are natural such as lighting strikes or humans leaving camp fires unattended. Which is true but wildfire can be caused by almost anything depending on region/area. Certain regions

depending on population, land and other factors are more likely to specific causes of wildfires. Wildfires are essentially fires that aren't a conrolled burn and are unplanned and mostly uncontrolled fires.

Reviewing other studies, it is found that certain areas/regions are prone to more wildfires. Such as grassland areas are more prone than forest areas are as the likelihood of a wildfire increases by almost twice in grassland (Lasslop et al 2017). Different ecosystems and types of land can help prevent or even start wildfires naturally. It is also seen and mentioned in other research that certain region like the midwest region of United States for example sees more human caused wildfires even though the midwest covers different land and soil types that may be more prone to natural wildfires (Cardille et al 2017). But a state in the west like California can see wildfires sparked by nature through lightning strikes which accounts for about 70% of the West region's wildfire causes (Syphard et al. 2024)

Some of the primary causes of wildfires can be due to nature especially in the west where it tends to be drier and space for wildfires to grow. But the coasts of the United States are dominated by human caused wildfires (Nagy et. al 2017). This can be almost correlated to the population as there is more individuals living near the coasts. Some research has investigated start points of wildfires and the distance from house, stores, and highways. It was also found regions/areas that were more populated were prone to wildfires but never found what exactly the cause could be other than human caused (Li et al. 2021).

The problem is to find out what is the primary cause of big wildfires in each US state, and when the states are grouped by region which causes are the most primary? This is essential to find out what are some of the primary causes of wildfires and find which regions the causes of wildfires are frequent/common. While the goal of this project is to also be specific with finding the causes of wildfire in each region and not to just look at if the wildfire was natural or human caused. But rather was it arson, firearms, fireworks, and etc.. Some controversies of this that were found is many wildfires that start never have a definite cause and an unknown origin as wildfires start up fast and can burn for long periods of time. This project hopes to answer the question of what is the primary cause of big wildfires in each US state, and when the states are grouped by region which causes are the most primary?

Methods

Data

The dataset was found on: https://www.kaggle.com/datasets/behroozsohrabi/us-wildfire-records-6th-edition/data. This dataset is for wildfires that happened from 1992 to 2020. Some of the things the dataset contains is: the year, discovery date of the fire, causes, size of fire, state, county, latitude, and longitude. Also, some state data was used: 'https://www2.census.gov/geo/tiger/GENZ2021/shp/cb_2021_us_state_500k.zip'. This data included: State fip code, state name, state abbreviations. The column name as follows: STATEFP, STATENS, AFFGEOID, GEOID, STUSPS, NAME, LSAD ALAND,AWATER

Data cleaning/munging

Many columns were dropped as they weren't needed for this project. Column names in the dataset were renamed for better usability. In the STATEFP column the state abreviations were turned into state FIP codes for easier access when merging. The biggest fire based on size in each state from 1992 to 2020 is taken and put into a new dataset called BigWF. groupby() is used. Each state was assigned an region according to the https://www.cdc.gov/nchs/hus/sources-definitions/geographic-region.htm. An column "Region" was created within the merged_gdf data: Northeast, Midwest, South or West being assigned. Then using groupby again the columns Region and Fire Size are used to find the biggest fire in each region. Then columns Region and General Cause is used to find the cause of the biggest fire in each region.

Data manipulation

The new dataset BigWF (biggest wildfires data) and state_gdf (state data) were merged together so it can be plotted. Both these datasets were merged under merged_gdf. Another, new data set was also created to help see if Cause Classification and General Cause can be used to predict Fire Size which was called results. A machine learning method used was Random Forest Classifier to try and predict fire size was unsuccessful. The random forest classifier was used to try and accurately predict the Fire Size based on just using the Cause Classification (Human, Nature) and General Cause. But it didn't come close to predicting any of the Fire Sizes.

Results and Discussion:

Exploratory Data Analysis

The original data from: https://www.kaggle.com/datasets/behroozsohrabi/us-wildfire-records-6th-edition/data has 1048575 rows with 39 columns. This dataset contains data from 1992-2020 regarding any details of wildfires that occurred. So, the size, year, cause of fire, state, the fip code, the lattitude and longitude of the location of the fire. A lot of the columns that were missing values were removed so there was no need to replace the missing values as the column as a whole wasn't necessary.

.info was used after the dataset was cleaned up of columns that were not needed. This helped see exactly have type of values are being dealt with and if any need to be changed or modified. The number of rows and columns with NULL values is displayed.But the dataset and types as is, is good enough for me and the project. This overall helped me progress in the project knowing the datatypes, null values and amount of rows.

In [17]:
```
WF.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 9 columns):
 #   Column                Non-Null Count    Dtype  
---  ------                --------------    -----  
 0   Reporting Unit Name   1048575 non-null  object 
 1   Fire Name             468957 non-null   object 
 2   Fire Year             1048575 non-null  int64  
 3   Cause Classification  1048575 non-null  object 
 4   General Cause         1048575 non-null  object 
 5   Fire Size             1048575 non-null  float64
 6   LATITUDE              1048575 non-null  float64
 7   LONGITUDE             1048575 non-null  float64
 8   STATEFP               1048433 non-null  object 
dtypes: float64(3), int64(1), object(5)
memory usage: 72.0+ MB
```

.describe was used after the dataset was cleaned up of columns that were not needed.To help see the count, mean, STD, min and more for the columns of the dataset. This helps kind of get a feel for the data and help me progress knowing this. It wasn't need in this scenario.

In [18]:
```
WF.describe()
```

Out[18]:

	Fire Year	Fire Size	LATITUDE	LONGITUDE
count	1.048575e+06	1.048575e+06	1.048575e+06	1.048575e+06
mean	2.000766e+03	8.471163e+01	3.712294e+01	-9.641812e+01
std	5.019711e+00	2.677566e+03	6.044017e+00	1.542262e+01
min	1.992000e+03	9.000000e-05	1.793972e+01	-1.639693e+02
25%	1.997000e+03	1.000000e-01	3.300514e+01	-1.104507e+02
50%	2.001000e+03	1.000000e+00	3.535376e+01	-9.330250e+01
75%	2.005000e+03	4.000000e+00	4.110100e+01	-8.261544e+01
max	2.009000e+03	6.069450e+05	6.984950e+01	-6.525694e+01

In [19]:
```
WF.size
```

Out[19]:
```
9437175
```

To see the amount of different causes and how many times the causes were recorded is interesting to see as well:

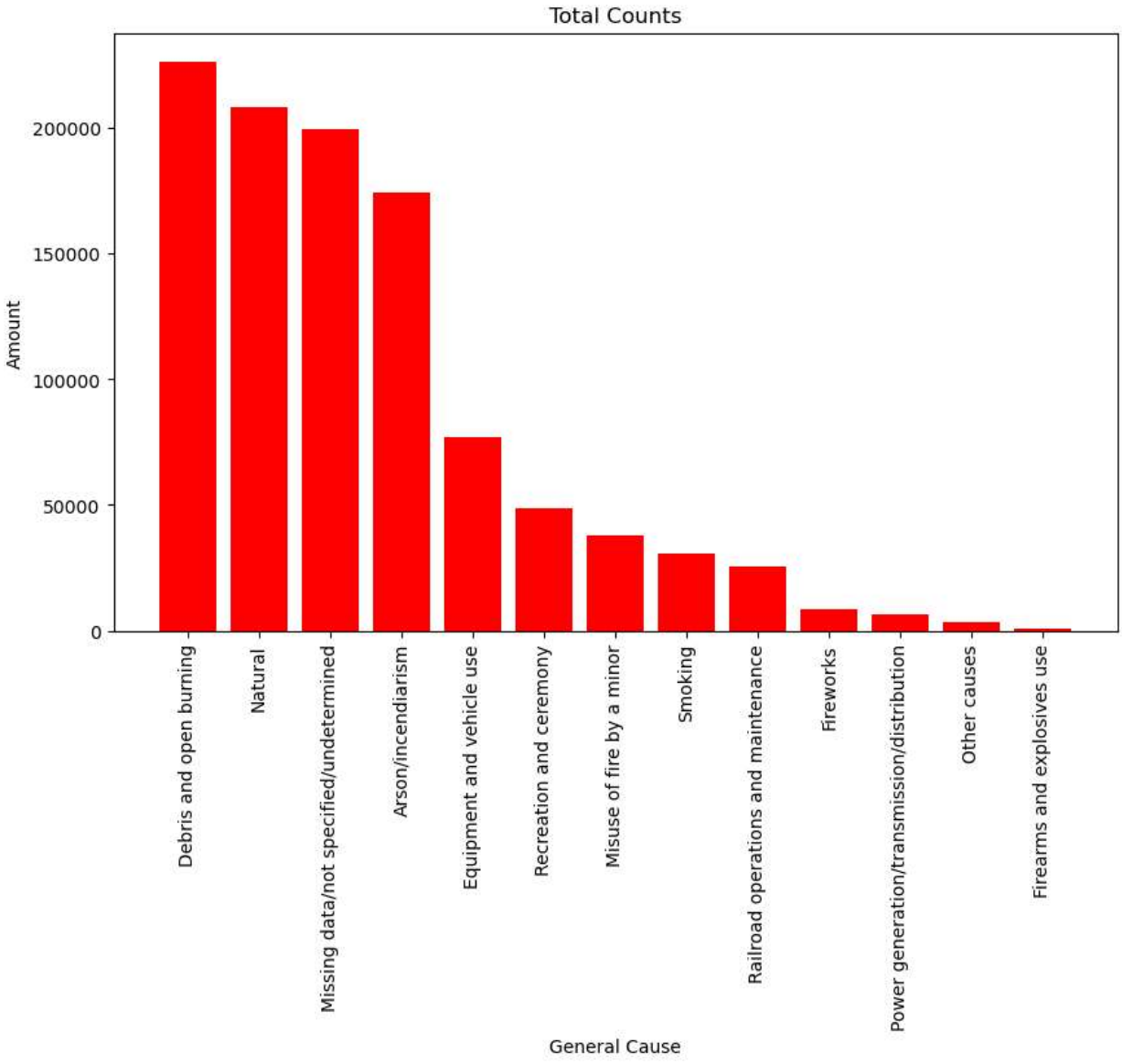
In [29]:
```
Count= WF['General Cause'].value_counts()
print(Count)
```

```
Debris and open burning                          226541
Natural                                          208411
Missing data/not specified/undetermined          199390
Arson/incendiarism                               174540
Equipment and vehicle use                         77196
Recreation and ceremony                           48729
Misuse of fire by a minor                         37759
Smoking                                           30855
Railroad operations and maintenance               25698
Fireworks                                          8710
Power generation/transmission/distribution         6642
Other causes                                       3530
Firearms and explosives use                         574
Name: General Cause, dtype: int64
```

Plots:

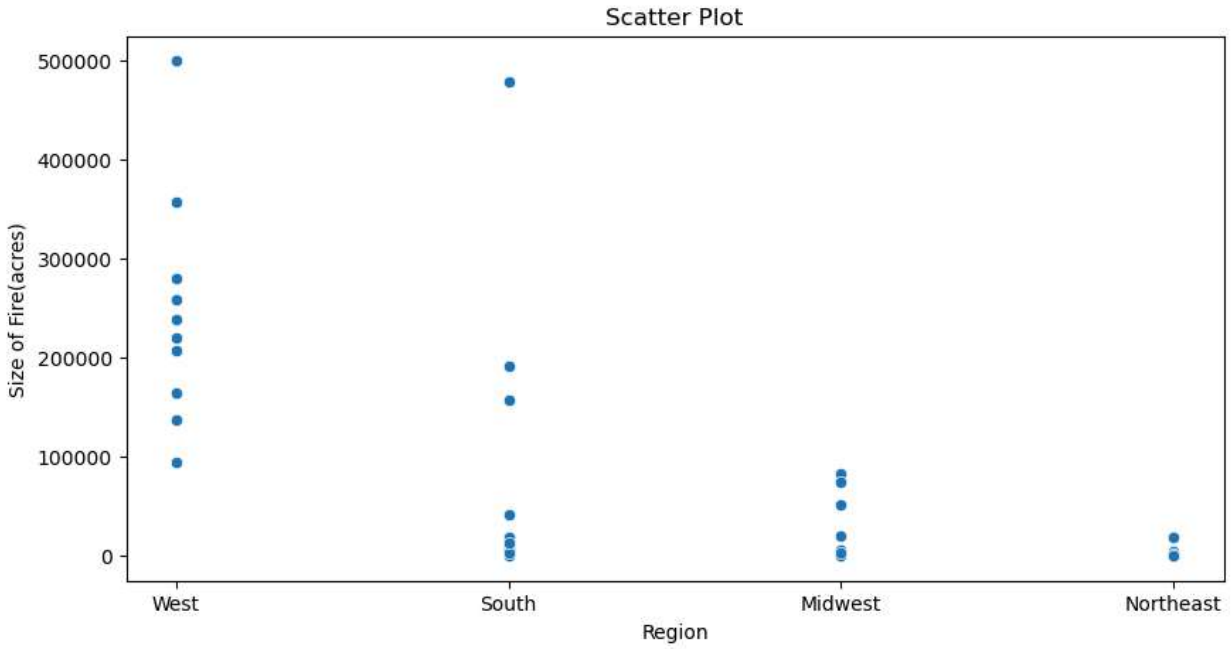
Below the bar graph helps visualize the cause of wildfires and the amount of those causes between 1992 and 2020. This kind of helps visualize that there is many causes of wildfires but the top 3 are debris and open burning, natural, and missing data/not specified/undetermined. Debris and open burning would be like camp fires being left unattended, natural would be like lightening strikes and missing data/not specified/undetermined would be the source of the wildfire was never determined.

In [12]:
```
Count= WF['General Cause'].value_counts()

plt.figure(figsize=(10, 6))
plt.bar(Count.index,Count.values,color= 'red')
plt.xlabel("General Cause")
plt.ylabel("Amount")
plt.title("Total Counts")
plt.xticks(rotation= 90)
plt.show()
```



Below the scatterplot helps us see the size of the wildfires in each region and its a bit interesting to look at because as you go west to east the sizes of the wildfires actually end up getting smaller. So a wildfire a person from the Northeast considers big is actually considered pretty small to a person from the west.

In [31]:
```
plt.figure(figsize= (10,5))
sns.scatterplot(x= 'Region',y= 'Fire Size',data= merged_gdf)
plt.title('Scatter Plot')
plt.xlabel('Region')
plt.ylabel('Size of Fire(acres)')

plt.show()
```



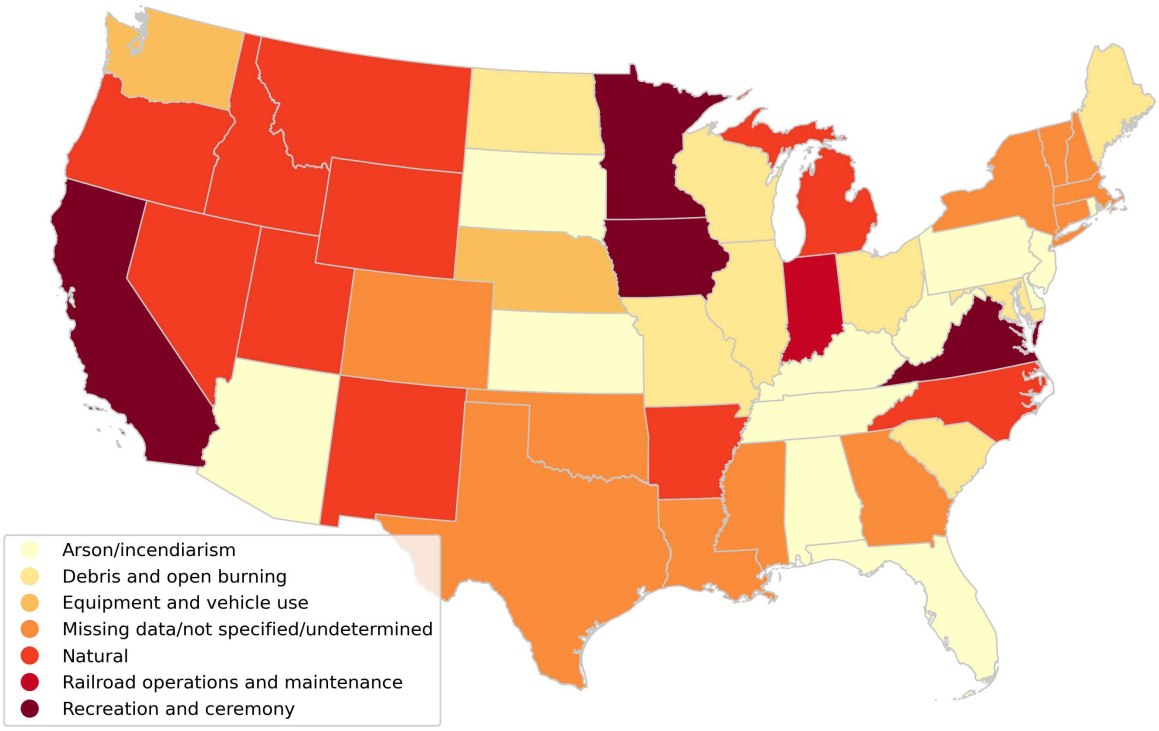
Below the table was made out of a curiosity aspect to see if a Random Forest classifier can accurately predict the Fire Size based on just using the Cause Classification(Human, Nature) and General Cause. Based on the table no it is not possible and would probably need a lot more to even get close to predicting the fire size.

In [13]:
```
results = pd.DataFrame({'Actual Fire Size': y_test, 'Predicted Fire Size': Pred})
print(results)
```

```
        Actual Fire Size  Predicted Fire Size
781974             50.00            37.031932
937737              0.30            31.651329
907828             10.00            31.651329
784628              2.00            36.733089
662460              2.00            36.733089
...                  ...                  ...
673443              1.50            31.651329
656736            100.00            37.031932
858501             27.00            31.651329
617079            854.00           274.270962
487559              0.15            31.651329

[209715 rows x 2 columns]
```

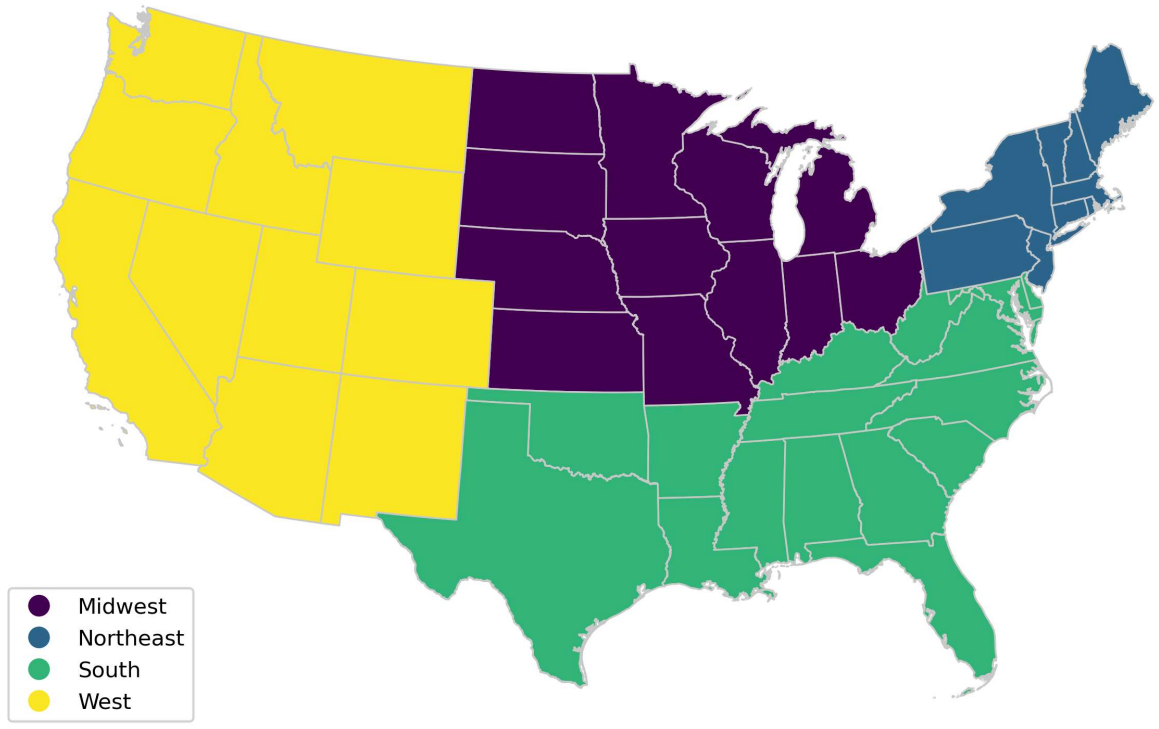
Cause of the biggest fire in each state 1992-2020. Below the map shows us a pretty diverse map of causes of the biggest fire in each state. Where natural is the main leader in cause of wildfires. While it is suprising and not suprising to see quite a bit of missing data/not specified/undetermined as there is many states that experienced there biggest wildfire and don't even know what exactly caused it.

In [20]:
```
merged_gdf= state_gdf.merge(BigWF,on= 'STATEFP')
fig,ax= plt.subplots(figsize= (12,8),dpi= 300)
merged_gdf.plot(column= 'General Cause',cmap= 'YlOrRd',linewidth= 0.8,ax= ax,edgecolor
                legend_kwds= {'loc':'lower left'})
ax.set_axis_off()
plt.show()
```



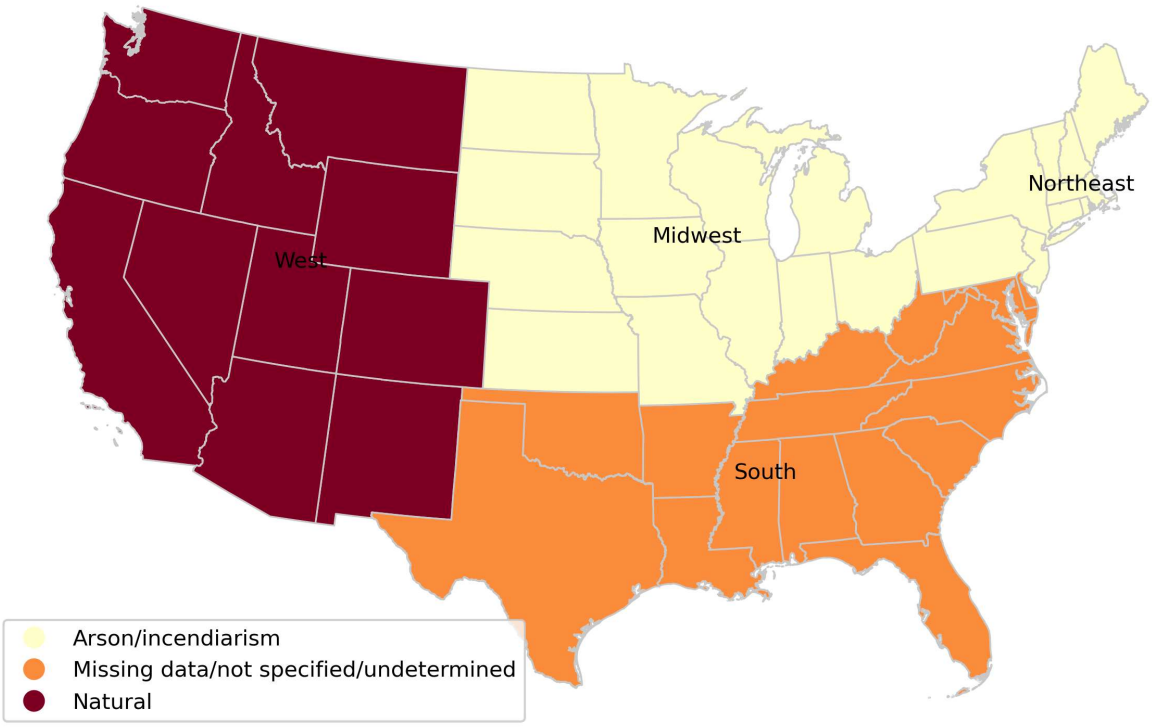
Every state is grouped by which region it belongs to below. This was done to help visualize which states belong to which regions.

In [24]:
```
fig,ax= plt.subplots(figsize= (10,8),dpi= 300)
merged_gdf.plot(column= 'Region',cmap= 'viridis',linewidth= 0.8,ax= ax,edgecolor= '0.8
                legend_kwds= {'loc':'lower left'})
ax.set_axis_off()
plt.show()
```



Below is the map and causes of the biggest fire in each Region 1992-2020. This helps get us to our conclusion in finding which cause started each regions biggest wildfire. Its a bit suprising to see again that missing data/not specified/undetermined is a big factor. As the south is missing data/not specified/undetermined, the west is natural causes and the midwest and northeast is arson.

In [27]:
```
fig,ax= plt.subplots(figsize= (10,8),dpi= 300)

merged_gdf.plot(column= 'BiggestFireCause',cmap= 'YlOrRd',linewidth= 0.8,ax= ax,edgecc
                legend_kwds= {'loc':'lower left'})
for region, data in merged_gdf.groupby('Region'):
    x,y= data.geometry.unary_union.centroid.xy
    ax.annotate(region,(x[0],y[0]))
ax.set_axis_off()
plt.show()
```



Conclusions

The main reason to find the biggest wildfire in each region and the cause of it was to show what cause has affected the region the most and how maybe more resources can be used to prevent a wildfire of the size to happen again. The take way from this project is that we now know what the cause of the biggest fire in each region is. While we know the causes for the West is natural and for the Midwest and Northeast it is arson. For the South we got the cause to be: missing data/not specified/undetermined. Which is a bit alarming as these fires were never either investigated into or all the evidence of what could have started it was burnt. It is not as surprising to see the cause for the West's biggest wildfire as the west region is a lot more open and drier so natural being the cause is fitting. While it is a lot more alarming to see that the

cause for the biggest wildfire in the Midwest and Northeast is arson. As this means this is more of a someone wanting to cause damage on people and properties. Now knowing the causes of the biggest wildfire in each region it could help in finding/placing resources to help prevent or lessen the effect of wildfires. If knowing the cause or at least assuming waht started a wildfire it can be helped to put out faster resulting in infrastructure and lives being saved. Based on the research done the east coast having its biggest cause of wildfire being arson is not surprising as we did learn more human made fires occur on the coasts. Same with the west the research was proved right as natural was the cause of the biggest wildfire because of the drier environment. The south having no answers was a bit surprising not to get an answer for. But, overall expected results were achieved.

Code Appendix:

In [1]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import geopandas as gpd
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
```

In [2]:
```
WF= pd.read_csv('WildFireData.csv',encoding= 'latin1',low_memory=False)
url_county= 'https://www2.census.gov/geo/tiger/GENZ2021/shp/cb_2021_us_state_500k.zip'
```

In [3]:
```
WF.head()
```

Out[3]:

	OBJECTID	Shape	FOD_ID	FPA_ID	SOURCE_SYSTEM_TYPE
0	1	b'\x00\x01\xad\x10\x00\x00\xc8\xce\n[_@^\xc0\x...	1	FS-1418826	FED
1	2	b'\x00\x01\xad\x10\x00\x00\xc8\xe594\xe2\x19^\...	2	FS-1418827	FED
2	3	b'\x00\x01\xad\x10\x00\x00x{\xac \x13/^\xc0@\x...	3	FS-1418835	FED
3	4	b'\x00\x01\xad\x10\x00\x00\xc8\x13u\xd7s\xfa]\...	4	FS-1418845	FED
4	5	b'\x00\x01\xad\x10\x00\x00\xd0\x11y\xf8\xb6\xf...	5	FS-1418847	FED

5 rows × 39 columns

In [4]:
```
state_gdf= gpd.read_file(url_county)
exclude_list= [15, 72, 2, 60, 66, 69, 78]
state_gdf= state_gdf.loc[~state_gdf['STATEFP'].astype(int).isin(exclude_list)]
state_gdf= state_gdf.to_crs(5070)
state_gdf.head()
```

Out[4]:

	STATEFP	STATENS	AFFGEOID	GEOID	STUSPS	NAME	LSAD	ALAND	AWATER	
0	56	01779807	0400000US56	56	WY	Wyoming	00	251458712294	1867503716	(
2	24	01714934	0400000US24	24	MD	Maryland	00	25151992308	6979074857	M
4	05	00068085	0400000US05	05	AR	Arkansas	00	134660767709	3121950081	
5	38	01779797	0400000US38	38	ND	North Dakota	00	178694310772	4414779956	
6	10	01779781	0400000US10	10	DE	Delaware	00	5046731559	1399179670	M

Many columns were dropped as they weren't need for this project

In [5]:
```
WF= WF.drop(['SOURCE_REPORTING_UNIT'],axis= 1)
WF= WF.drop(['LOCAL_FIRE_REPORT_ID'],axis= 1)
WF= WF.drop(['FIRE_CODE'],axis= 1)
WF= WF.drop(['ICS_209_PLUS_INCIDENT_JOIN_ID'],axis= 1)
WF= WF.drop(['ICS_209_PLUS_COMPLEX_JOIN_ID'],axis= 1)
WF= WF.drop(['MTBS_ID'],axis= 1)
WF= WF.drop(['MTBS_FIRE_NAME'],axis= 1)
WF= WF.drop(['COMPLEX_NAME'],axis= 1)
WF= WF.drop(['DISCOVERY_DATE'],axis= 1)
WF= WF.drop(['DISCOVERY_DOY'],axis= 1)
WF= WF.drop(['DISCOVERY_TIME'],axis= 1)
WF= WF.drop(['CONT_DATE'],axis= 1)
WF= WF.drop(['CONT_DOY'],axis= 1)
WF= WF.drop(['CONT_TIME'],axis= 1)
WF= WF.drop(['FIRE_SIZE_CLASS'],axis= 1)
WF= WF.drop(['OWNER_DESCR'],axis= 1)
WF= WF.drop(['NWCG_REPORTING_UNIT_NAME'],axis= 1)
WF= WF.drop(['NWCG_REPORTING_UNIT_ID'],axis= 1)
WF= WF.drop(['NWCG_REPORTING_AGENCY'],axis= 1)
WF= WF.drop(['SOURCE_SYSTEM'],axis= 1)
WF= WF.drop(['SOURCE_SYSTEM_TYPE'],axis= 1)
WF= WF.drop(['FPA_ID'],axis= 1)
WF= WF.drop(['FOD_ID'],axis= 1)
WF= WF.drop(['Shape'],axis= 1)
WF= WF.drop(['NWCG_CAUSE_AGE_CATEGORY'],axis= 1)
WF= WF.drop(['OBJECTID'],axis= 1)
WF= WF.drop(['LOCAL_INCIDENT_ID'],axis= 1)
WF= WF.drop(['FIPS_NAME'],axis= 1)
WF= WF.drop(['FIPS_CODE'],axis= 1)
WF= WF.drop(['COUNTY'],axis= 1)
```

Column names in the dataset were renamed here for better usability.

In [6]:
```
ColumnName= {'NWCG_CAUSE_CLASSIFICATION': 'Cause Classification','NWCG_GENERAL_CAUSE':
             'FIPS_NAME':'County Name','FIRE_SIZE': 'Fire Size','FIRE_YEAR':'Fire Year'
              'SOURCE_REPORTING_UNIT_NAME':'Reporting Unit Name','COUNTY':'County FIPS'
WF.rename(columns=ColumnName, inplace=True)
```

In [28]:
```
WF.head()
```

Out[28]:

	Reporting Unit Name	Fire Name	Fire Year	Cause Classification	General Cause	Fire Size	LATITUDE	LON
0	Plumas National Forest	FOUNTAIN	2005	Human	Power generation/transmission/distribution	0.10	40.036944	-12
1	Eldorado National Forest	PIGEON	2004	Natural	Natural	0.25	38.933056	-12(
2	Eldorado National Forest	SLACK	2004	Human	Debris and open burning	0.10	38.984167	-12(
3	Eldorado National Forest	DEER	2004	Natural	Natural	0.10	38.559167	-11
4	Eldorado National Forest	STEVENOT	2004	Natural	Natural	0.10	38.559167	-11

New data set created to help see if Cause Classification and General Cause can be used to predict Fire Size. Its a long shot using Random Forest classifier but sometimes similar causes of wildfires burn similar and cause damage that way so it could be interesting if possible to predict this way.

In [7]:
```
WF_new= WF[['Cause Classification', 'Fire Size','General Cause']]
WF_new1= pd.get_dummies(WF_new, columns=['Cause Classification','General Cause'])
```

Random Forest classifier below:

In [8]:
```
X= WF_new1.drop('Fire Size', axis=1)
y= WF_new1['Fire Size']

X_train,X_test,y_train,y_test= train_test_split(X,y,test_size= 0.2,random_state= 42)
Regr= RandomForestRegressor(n_estimators= 100,max_depth= 2,random_state= 0)
Regr.fit(X_train,y_train)
Pred= Regr.predict(X_test)
```

Based on results below, Fire Size can't be predicted just based on Cause Classification and General Cause

In [9]:
```
results= pd.DataFrame({'Actual Fire Size': y_test, 'Predicted Fire Size': Pred})
print(results)
```

```
        Actual Fire Size  Predicted Fire Size
781974             50.00            37.031932
937737              0.30            31.651329
907828             10.00            31.651329
784628              2.00            36.733089
662460              2.00            36.733089
...                  ...                  ...
673443              1.50            31.651329
656736            100.00            37.031932
858501             27.00            31.651329
617079            854.00           274.270962
487559              0.15            31.651329

[209715 rows x 2 columns]
```

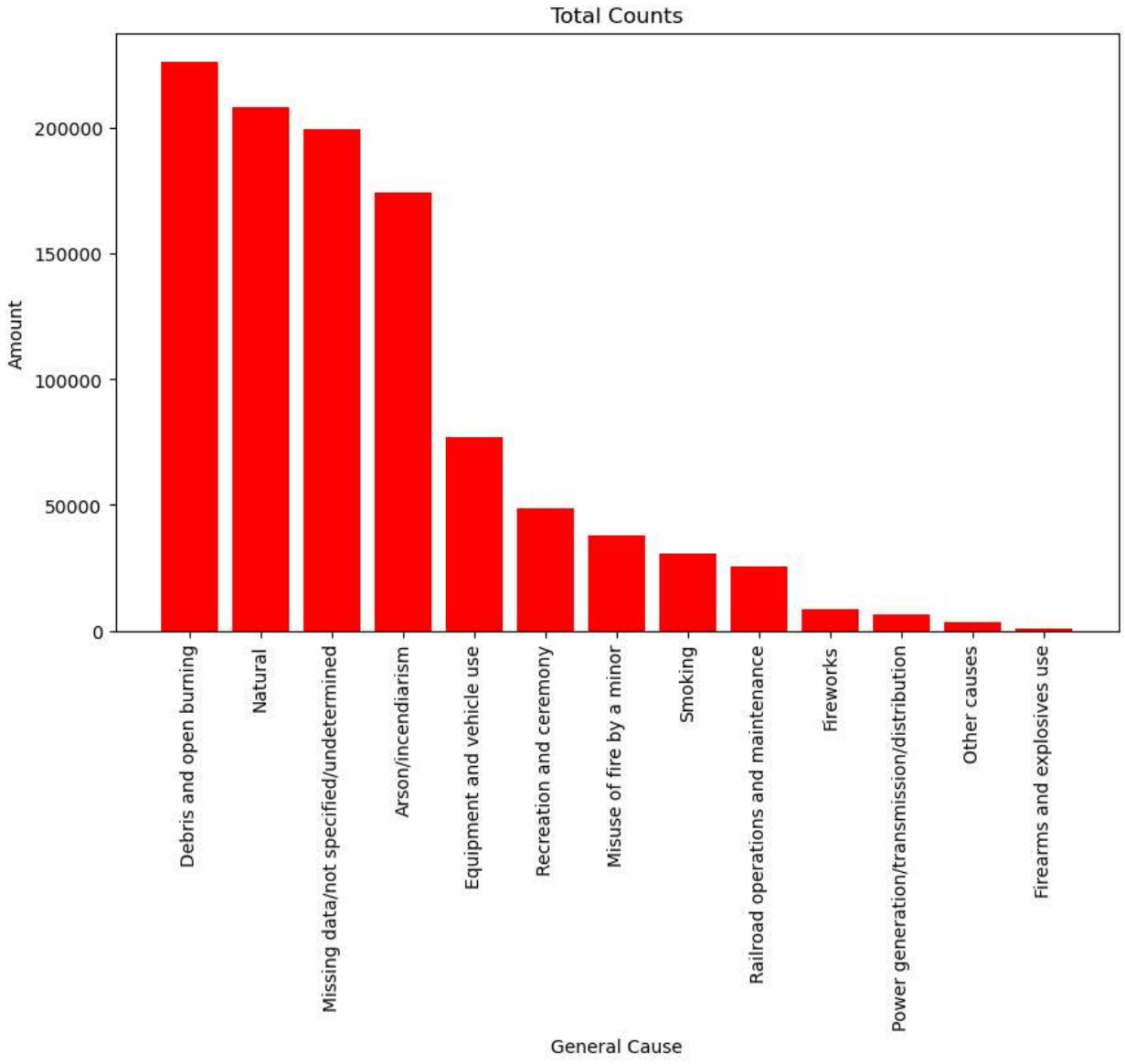
Column count for General Causes to help show the amount of each cause of wildfire.

In [10]:
```
Count= WF['General Cause'].value_counts()
print(Count)
```

```
Debris and open burning                      226541
Natural                                      208411
Missing data/not specified/undetermined      199390
Arson/incendiarism                           174540
Equipment and vehicle use                     77196
Recreation and ceremony                       48729
Misuse of fire by a minor                     37759
Smoking                                       30855
Railroad operations and maintenance           25698
Fireworks                                      8710
Power generation/transmission/distribution     6642
Other causes                                   3530
Firearms and explosives use                     574
Name: General Cause, dtype: int64
```

Bar plot to help visualize the amount of each cause of wildfire.

In [11]:
```
Count= WF['General Cause'].value_counts()

plt.figure(figsize=(10, 6))
plt.bar(Count.index,Count.values,color= 'red')
plt.xlabel("General Cause")
plt.ylabel("Amount")
plt.title("Total Counts")
plt.xticks(rotation= 90)
plt.show()
```



In the STATEFP column the state abreviations were turned into state FIP codes for easier use

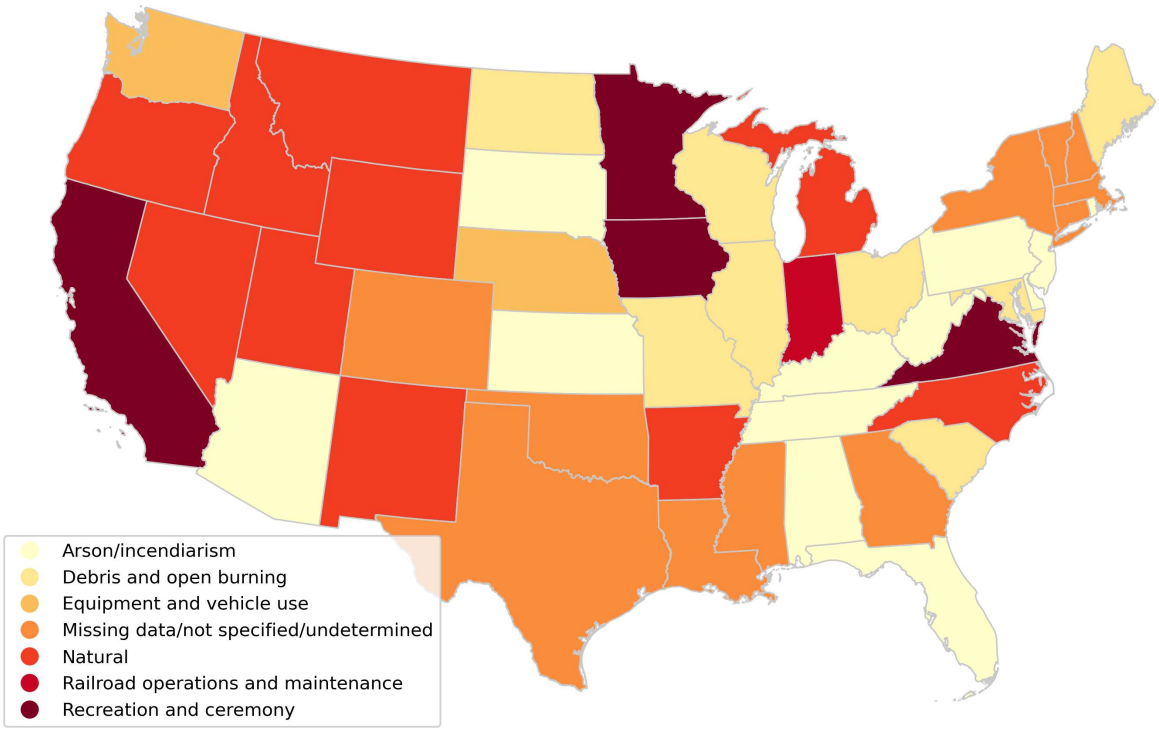
In [14]:
```
STATE= {
    'AL':'01','AK':'02','AZ':'04','AR':'05','CA':'06','CO':'08','CT':'09','DE':'10','D
    'GA':'13','HI':'15','ID':'16','IL':'17','IN':'18','IA':'19','KS':'20','KY':'21','L
    'MD':'24','MA':'25','MI':'26','MN':'27','MS':'28','MO':'29','MT':'30','NE':'31','N
    'NJ':'34','NM':'35','NY':'36','NC':'37','ND':'38','OH':'39','OK':'40','OR':'41','P
    'SC':'45','SD':'46','TN':'47','TX':'48','UT':'49','VT':'50','VA':'51','WA':'53','W
}
WF['STATEFP'] = WF['STATEFP'].map(STATE)
```

The biggest fire based on size in each state from 1992 to 2020 is taken and put into a new dataset. groupby was used here.

In [15]:
```
BigWF= WF.loc[WF.groupby('STATEFP')['Fire Size'].idxmax()]
pd.set_option('display.max_columns', None)
```

Cause of the biggest fire in each state plot 1992-2020

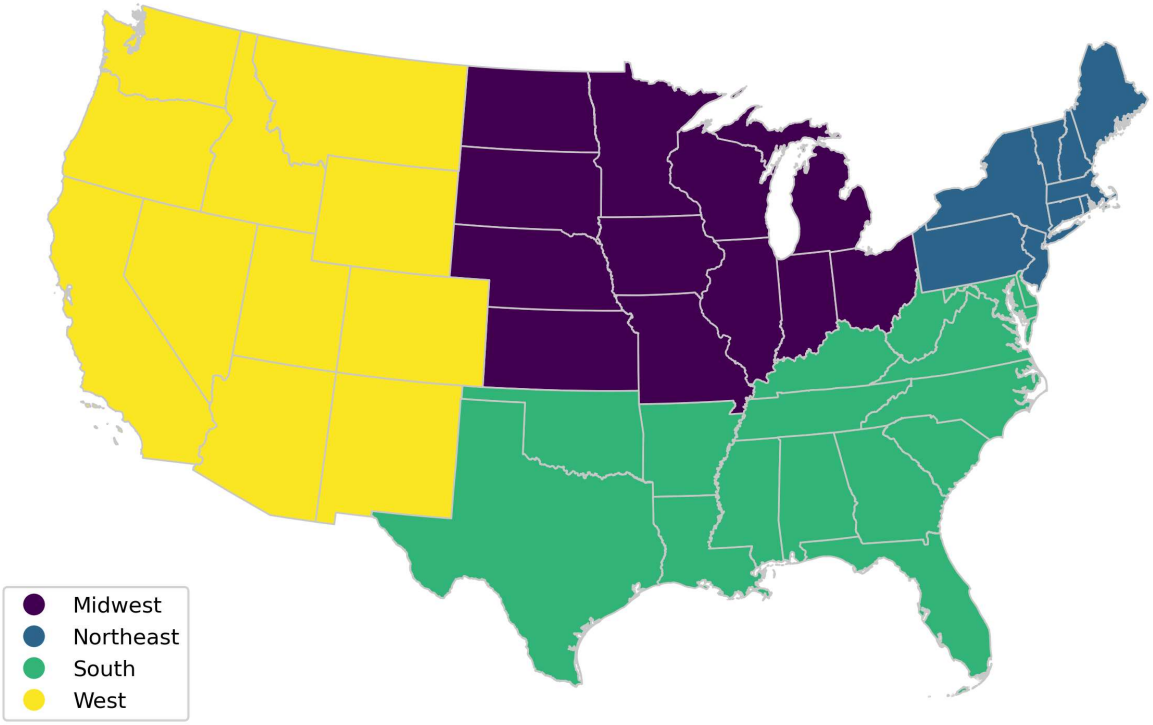
In [16]:
```
merged_gdf= state_gdf.merge(BigWF,on= 'STATEFP')
fig,ax= plt.subplots(figsize= (12,8),dpi= 300)
merged_gdf.plot(column= 'General Cause',cmap= 'YlOrRd',linewidth= 0.8,ax= ax,edgecolor
                legend_kwds= {'loc':'lower left'})
ax.set_axis_off()
plt.show()
```



In [22]:
```
regions = {
    "Northeast": ["09", "23", "25", "33", "34", "36", "42", "44", "50"],
    "Midwest": ["17", "18", "19", "20", "26", "27", "29", "31", "38", "39", "46", "55"
    "South": ["01", "05", "10", "11", "12", "13", "21", "22", "24", "28", "37", "40",
    "West": ["02", "04", "06", "08", "15", "16", "30", "32", "35", "41", "49", "53", "
}

SR= {state: region for region, states in regions.items() for state in states}
merged_gdf['Region']= merged_gdf['STATEFP'].map(SR)
```

Every State is grouped by which region it belongs to below:

In [23]:
```
fig,ax= plt.subplots(figsize= (10,8),dpi= 300)
merged_gdf.plot(column= 'Region',cmap= 'viridis',linewidth= 0.8,ax= ax,edgecolor= '0.8
                legend_kwds= {'loc':'lower left'})
ax.set_axis_off()
plt.show()
```

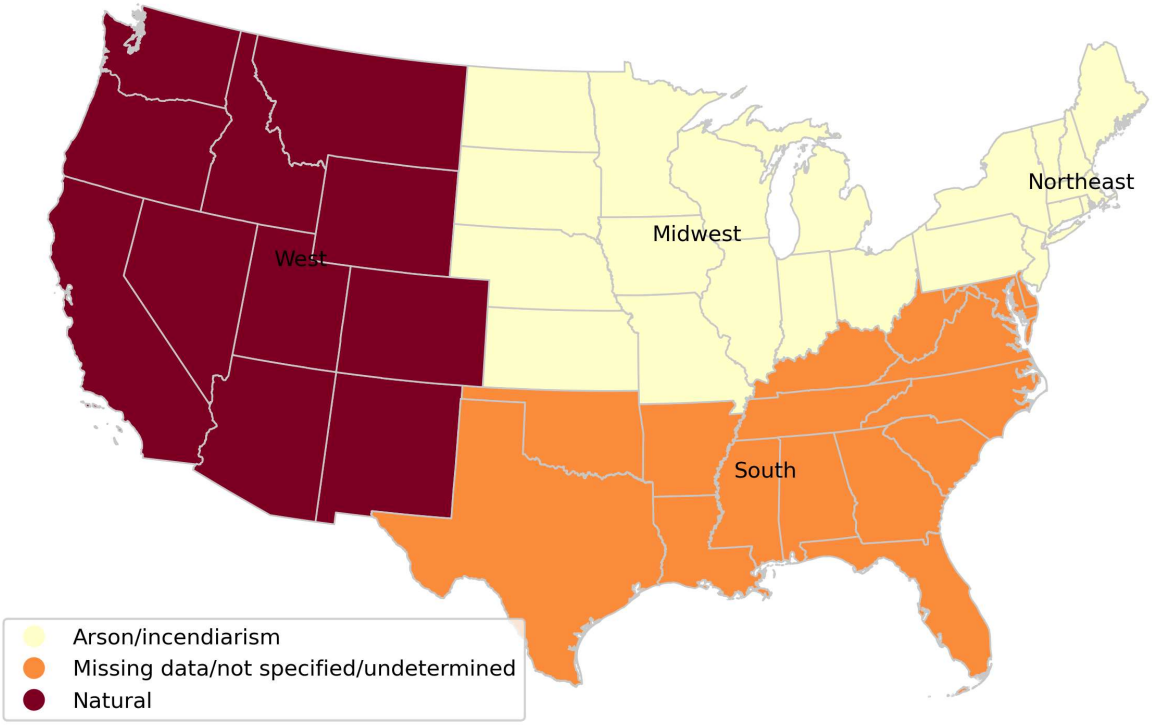


Using groupby the columns Region and Fire Size are used to find the biggest fire in each region. Then columns Region and General Cause is used to find the the cause of the biggest fire in each region. groupby is used here as well.

In [25]:
```
BF= merged_gdf.groupby('Region')['Fire Size'].idxmax()
BRRegiond= merged_gdf.loc[BF]
RC= BRRegiond.set_index('Region')['General Cause'].to_dict()
merged_gdf['BiggestFireCause']= merged_gdf['Region'].map(RC)
```

Below is the plot for the cause of the biggest fire in each Region 1992-2020.

In [26]:
```
fig,ax= plt.subplots(figsize= (10,8),dpi= 300)

merged_gdf.plot(column= 'BiggestFireCause',cmap= 'YlOrRd',linewidth= 0.8,ax= ax,edgecc
                legend_kwds= {'loc':'lower left'})
for region, data in merged_gdf.groupby('Region'):
    x,y= data.geometry.unary_union.centroid.xy
    ax.annotate(region,(x[0],y[0]))
ax.set_axis_off()
plt.show()
```



Bibliography

1. Lasslop, Gitta, and Silvia Kloster. "Human impact on wildfires varies between regions and with vegetation productivity." Environmental Research Letters 12.11 (2017): 115011.
2. Cardille, Jeffrey A., Stephen J. Ventura, and Monica G. Turner. "Environmental and social factors influencing wildfires in the Upper Midwest, United States." Ecological applications 11.1 (2017): 111-127.
3. Syphard, Alexandra D., and Jon E. Keeley. "Location, timing and extent of wildfire vary by cause of ignition." International Journal of Wildland Fire 24.1 (2024)
4. Syphard, Alexandra D., and Jon E. Keeley. "Location, timing and extent of wildfire vary by cause of ignition." International Journal of Wildland Fire 24.1 (2015): 37-47.
5. Li, Shu, and Tirtha Banerjee. "Spatial and temporal pattern of wildfires in California from 2000 to 2019." Scientific reports 11.1 (2021): 8779.