Wildfire Data Preprocessing

[Link to Notebook]

```
import pandas as pd
import numpy as np
import requests

from io import StringIO

from datetime import datetime

!pip install geopandas
import geopandas

import pandas as pd
import numpy as np
from math import radians, cos, sin, asin, sqrt
import requests
from io import StringIO


# Important library for many geopython libraries
!apt install gdal-bin python-gdal python3-gdal
# Install rtree - Geopandas requirment
!apt install python3-rtree
# Install Geopandas
!pip install git+git://github.com/geopandas/geopandas.git
# Install descartes - Geopandas requirment
!pip install descartes
# Install Folium for Geographic data visualization
!pip install folium
# Install plotlyExpress
!pip install plotly_express

# ca bioregions shape file
import geopandas as gpd

!apt install libspatialindex-dev
!pip3 install rtree
!pip3 install pygeos
```

```
  Running command git clone -q git://github.com/geopandas/geopandas.git /tmp/p
Requirement already satisfied (use --upgrade to upgrade): geopandas==0.8.0+70.
Requirement already satisfied: pandas>=0.24.0 in /usr/local/lib/python3.6/dist
Requirement already satisfied: shapely>=1.6 in /usr/local/lib/python3.6/dist-p
Requirement already satisfied: fiona>=1.8 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: pyproj>=2.2.0 in /usr/local/lib/python3.6/dist-
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.6/dist-
```

```
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.6/dist-p
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python
Requirement already satisfied: cligj>=0.5 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: click<8,>=4.0 in /usr/local/lib/python3.6/dist-
Requirement already satisfied: click-plugins>=1.0 in /usr/local/lib/python3.6/
Requirement already satisfied: munch in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: attrs>=17 in /usr/local/lib/python3.6/dist-pack
Requirement already satisfied: certifi in /usr/local/lib/python3.6/dist-packag
Requirement already satisfied: six>=1.7 in /usr/local/lib/python3.6/dist-packa
Building wheels for collected packages: geopandas
  Building wheel for geopandas (setup.py) ... done
  Created wheel for geopandas: filename=geopandas-0.8.0+70.gc325570-py2.py3-no
  Stored in directory: /tmp/pip-ephem-wheel-cache-12xai9e4/wheels/91/24/71/376
Successfully built geopandas
Requirement already satisfied: descartes in /usr/local/lib/python3.6/dist-pack
Requirement already satisfied: matplotlib in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: numpy>=1.11 in /usr/local/lib/python3.6/dist-pa
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /us
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.6/d
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.6/dist-p
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.6/dist-packa
Requirement already satisfied: folium in /usr/local/lib/python3.6/dist-package
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: requests in /usr/local/lib/python3.6/dist-packa

Requirement already satisfied: jinja2 in /usr/local/lib/python3.6/dist-package
Requirement already satisfied: six in /usr/local/lib/python3.6/dist-packages (
Requirement already satisfied: branca>=0.3.0 in /usr/local/lib/python3.6/dist-
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.6/d
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.6/dist-p
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.6/di
Requirement already satisfied: plotly_express in /usr/local/lib/python3.6/dist
Requirement already satisfied: numpy>=1.11 in /usr/local/lib/python3.6/dist-pa
Requirement already satisfied: patsy>=0.5 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: statsmodels>=0.9.0 in /usr/local/lib/python3.6/
Requirement already satisfied: scipy>=0.18 in /usr/local/lib/python3.6/dist-pa
Requirement already satisfied: pandas>=0.20.0 in /usr/local/lib/python3.6/dist
Requirement already satisfied: plotly>=4.1.0 in /usr/local/lib/python3.6/dist-
Requirement already satisfied: six in /usr/local/lib/python3.6/dist-packages (
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.6/dist-p
Requirement already satisfied: retrying>=1.3.3 in /usr/local/lib/python3.6/dis
Reading package lists... Done
Building dependency tree
Reading state information... Done
libspatialindex-dev is already the newest version (1.8.5-5).
0 upgraded, 0 newly installed, 0 to remove and 14 not upgraded.
Requirement already satisfied: rtree in /usr/lib/python3/dist-packages (0.8.3)
Requirement already satisfied: pygeos in /usr/local/lib/python3.6/dist-package
```

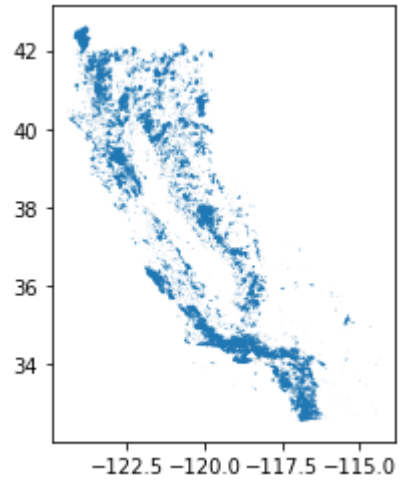
Load Data

```
shapeFile = geopandas.read_file("https://opendata.arcgis.com/datasets/e3802d2abf8741a
```

```
shapeFile = shapeFile.to_crs("EPSG:4326")
```

```
shapeFile["center"] = shapeFile["geometry"].centroid
shapeFile.plot()
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: UserWarning: Geo

  """Entry point for launching an IPython kernel.
<matplotlib.axes._subplots.AxesSubplot at 0x7fb7d26aa748>
```



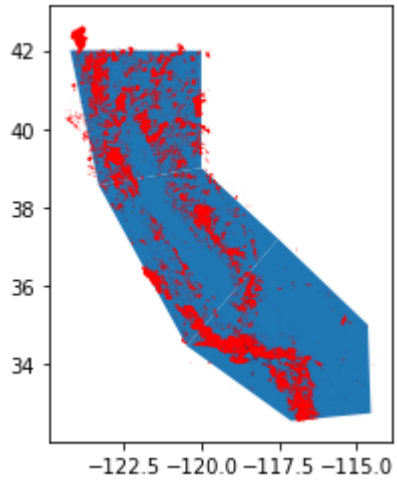
Fit into Bioregions

```
import geopandas as gpd
from shapely.geometry import Polygon

poly1_lons = [41.996221, 41.985584,38.99915, 38.556104]
poly1_lats = [-124.216027, -120.003800, -120.00038, -123.298092]
poly2_lons = [38.556104, 38.99915, 37.190380, 34.463602]
poly2_lats = [-123.298092, -120.00038, -117.528050, -120.46687]
reg1 = Polygon(zip(poly1_lats, poly1_lons))
reg2 = Polygon(zip(poly2_lats, poly2_lons))
reg3 = Polygon([(-117.528050,37.190380), (-120.46687,34.463602), (-117.116231,32.5524

df_bioregions = gpd.GeoDataFrame(geometry=[reg1, reg2,reg3])
df_bioregions['region'] = [1,2,3]
```

```
import matplotlib.pyplot as plt


fig, ax = plt.subplots()
```

```
df_bioregions.plot(ax=ax)
shapeFile.plot(ax=ax, color="red")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb7d2676a20>
```




```
import rtree
import pygeos

join = geopandas.sjoin(df_bioregions, shapeFile, how="inner", op="intersects")
join
```

0	POLYGON ((-124.21603 41.99622, -120.00380 41.9...	1	9045	9046	1978	CA	CDF	SNU	CREI(
1	POLYGON ((-123.29809 38.55610, -120.00038 38.9...	2	9045	9046	1978	CA	CDF	SNU	CREI(
0	POLYGON ((-124.21603 41.99622, -120.00380 41.9...	1	8825	8826	1954	CA	CDF	SNU	CH
1	POLYGON ((-123.29809 38.55610, -120.00038 38.9...	2	8825	8826	1954	CA	CDF	SNU	CH
0	POLYGON ((-124.21603 41.99622, -120.00380 41.9...	1	8739	8740	1953	CA	CDF	LNU	CH
...	...	...	...	...	...	...	...	...	
2	POLYGON ((-117.52805 37.19038, -120.46687 34.4...	3	18597	18604	2006	CA	NPS	DVP	S
2	POLYGON ((-117.52805 37.19038, -120.46687 34.4...	3	18617	18624	2012	CA	NPS	DVP	
2	POLYGON ((-117.52805 37.19038,	3	13815	13816	1971	CA	USF	INF	

```
join = join.dropna(subset=["ALARM_DATE"])

join["StartTime"] = join.apply(lambda x: datetime.strptime(str(x["ALARM_DATE"]), "%Y-
join["Month"] = join.apply(lambda x: x["StartTime"].month, axis=1)
join = join[shapeFile["YEAR_"] >= "1970"]
join
```

	-120.00380 41.9...								
1	POLYGON ((-123.29809 38.55610, -120.00038 38.9...	2	9045	9046	1978	CA	CDF	SNU	CREI(
0	POLYGON ((-124.21603 41.99622, -120.00380 41.9...	1	8825	8826	1954	CA	CDF	SNU	CH
1	POLYGON ((-123.29809 38.55610, -120.00038 38.9...	2	8825	8826	1954	CA	CDF	SNU	CH
0	POLYGON ((-124.21603 41.99622, -120.00380 41.9...	1	19709	20335	2017	CA	CDF	LNU	
...	...	...	...	...	...	...	...	...	
2	POLYGON ((-117.52805 37.19038, -120.46687 34.4...	3	18597	18604	2006	CA	NPS	DVP	S
2	POLYGON ((-117.52805 37.19038, -120.46687 34.4...	3	18617	18624	2012	CA	NPS	DVP	
2	POLYGON ((-117.52805 37.19038, -120.46687 34.4...	3	13815	13816	1971	CA	USF	INF	
2	POLYGON ((-117.52805 37.19038, -120.46687 34.4...	3	17513	17514	2013	CA	CDF	BDU	
2	POLYGON ((-117.52805 37.19038, -120.46687 34.4...	3	18651	18658	2013	CA	NPS	DVP	S(

15103 rows × 25 columns

```
pd.DataFrame(join.groupby(["region", "YEAR_", "Month"])["GIS_ACRES"].sum()).to_csv("a
```

```
pd.DataFrame(join.groupby(["region", "YEAR_", "Month"])["OBJECTID"].count()).to_csv("
```

```
shapeFile
```

1	2	2007	CA	CCO	LAC	MAGIC	00233077	2007-10-22	2
2	3	2007	CA	USF	ANF	RANCH	00000166	2007-10-20	2
3	4	2007	CA	CCO	LAC	EMMA	00201384	2007-09-11	2
4	5	2007	CA	CCO	LAC	CORRAL	00259483	2007-11-24	2
...	...	...	...	...	...	...	...	...	
20809	21435	2019	CA	CCO	LAC	MUREAU	None	2019-10-30	2
20810	21436	2019	CA	LRA	None	OAK	None	2019-10-28	2
20811	21437	2019	CA	LRA	LDF	BARHAM	00000845	2019-11-09	2
20812	21438	2019	CA	NPS	MNP	STAR	00013598	None	
20813	21439	2019	CA	LRA	LDF	SADDLE RIDGE	00001582	2019-10-10	2