Wildfire_Analysis_LocalPaths-checkpoint-checkpoint-checkpoint

June 19, 2025

1 Wildfire Analysis: Fire Area vs. Distance to Nearest Fire Station

This notebook analyzes fire perimeters in San Diego County to test the hypothesis: *Fires farther from fire stations tend to burn more area.*

[2]:
```
import geopandas as gpd
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from shapely.geometry import shape
import fiona
from scipy.stats import pearsonr
```

[41]:
```
# Step 1: Set and project fire perimeters first
fire_perimeters = gpd.read_file('California_Fire_Perimeters_2017.shp')

# Only assign CRS if it's missing
if fire_perimeters.crs is None:
    fire_perimeters = fire_perimeters.set_crs('EPSG:4326')

# Project perimeters to UTM 11N for San Diego
fire_perimeters_proj = fire_perimeters.to_crs('EPSG:32611')

# Step 2: Calculate centroids and area (km²) from projected perimeters
fire_perimeters_proj['geometry_centroid'] = fire_perimeters_proj.geometry.
 ↪centroid
fire_perimeters_proj['area_km2'] = fire_perimeters_proj.geometry.area / 1e6

# Step 3: Build the fire_centroids_proj GeoDataFrame directly
fire_centroids_proj = gpd.GeoDataFrame({
    'FIRE_NAME': fire_perimeters_proj['FIRE_NAME'],
    'YEAR_': fire_perimeters_proj['YEAR_'],
    'area_km2': fire_perimeters_proj['area_km2'],
    'geometry': fire_perimeters_proj['geometry_centroid']
}, crs='EPSG:32611')
```

[4]:
```
fire_centroids = gpd.GeoDataFrame({
    'FIRE_NAME': fire_names,
    'YEAR_': years,
    'area_km2': areas_km2,
    'geometry': centroids
}, crs='EPSG:4326')
```

[51]:
```
# Reproject fire stations
fire_stations.set_crs('EPSG:4326', inplace=True)
fire_stations_proj = fire_stations.to_crs('EPSG:32611')

#  Final clean recalculation after all CRS/geometry fixes
fire_centroids_proj['nearest_dist_km'] = fire_centroids_proj.geometry.apply(
    lambda pt: fire_stations_proj.distance(pt).min() / 1000  # meters to km
)
```

[53]:
```
analysis_df = fire_centroids_proj[['FIRE_NAME', 'YEAR_', 'area_km2',␣
 ↪'nearest_dist_km']]
corr, p_value = pearsonr(analysis_df['nearest_dist_km'],␣
 ↪analysis_df['area_km2'])
```

[25]:
```
fire_stations_proj.is_valid.value_counts()
```

[25]:
```
True    307
Name: count, dtype: int64
```

[27]:
```
fire_stations_proj.head(3)
```

[27]:
```
   objectid                       stat_name stat_type seed dist_name juris  \
0         1   USFS Oak Grove Fire Station 31  Seasonal    Y      USFS    CN
1         2     USFS Palomar Fire Station 36  Seasonal    Y      USFS    CN
2         3  USFS Cottonwood Fire Station 44  Seasonal    Y      USFS    CN

  dispatch       phone_num  sta_num submappage sdfdpggrid designator  \
0      CNF  (619) 767-9744       31    7843-B2       None        CNF
1      CNF  (760) 742-3491       36    7639-A2       None        CNF
2      CNF  (619) 473-9835       44    2058-B2       None        CNF

  assets_ava                                          address symbol  \
0        EMT   37560 Highway 79, Warner Springs, CA, 92086, USA    FSL
1        EMT                                             None    FSL
2        EMT  3971 Buckman Springs Rd, Pine Valley, CA, 9196…    FSL

   battalion                   geometry
0          5  POINT (5.19e+05 3.69e+06)
1          7  POINT (5.12e+05 3.69e+06)
2          4  POINT (5.48e+05 3.62e+06)
```

[29]:
```
print(fire_stations.crs)
print(fire_stations_proj.crs)
```

```
EPSG:4326
EPSG:32611
```

[31]:
```
fire_centroids_proj.is_valid.value_counts()
```

[31]:
```
False    608
Name: count, dtype: int64
```

[33]:
```
fire_centroids.is_valid.value_counts()
```

[33]:
```
True    608
Name: count, dtype: int64
```

[37]:
```
fire_centroids_proj.is_valid.value_counts()
```

[37]:
```
False    608
Name: count, dtype: int64
```

[43]:
```
fire_centroids_proj.is_valid.value_counts()
```

[43]:
```
True    608
Name: count, dtype: int64
```

[13]:
```
len(fire_centroids_proj)
```

[13]:
```
608
```

[15]:
```
fire_centroids_proj['nearest_dist_km'].value_counts()
```

[15]:
```
nearest_dist_km
1.797693e+305    608
Name: count, dtype: int64
```

[17]:
```
fire_centroids_proj['area_km2'].describe()
```

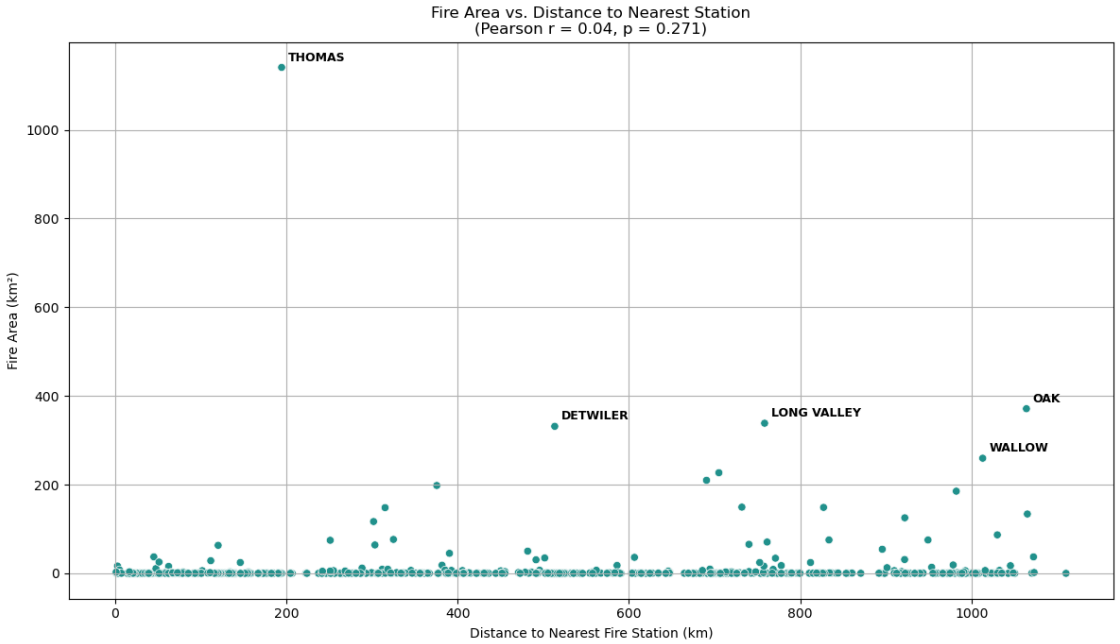
[17]:
```
count     608.000000
mean       15.373824
std        89.832058
min         0.000008
25%         0.057752
50%         0.185527
75%         0.910724
max      1681.105895
Name: area_km2, dtype: float64
```

[45]:
```
top_area = analysis_df.nlargest(5, 'area_km2')
top_dist = analysis_df.nlargest(5, 'nearest_dist_km')
highlight_fires = pd.concat([top_area, top_dist]).drop_duplicates()
```

[57]:
```
plt.close('all')
```

[63]:
```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(12, 7))
ax = sns.scatterplot(
    data=analysis_df,
    x='nearest_dist_km',
    y='area_km2',
    hue='YEAR_',
    palette='viridis',
    legend=False
)

# Add labels for large fires
for _, row in highlight_fires.iterrows():
    ax.annotate(
        row['FIRE_NAME'],
        (row['nearest_dist_km'], row['area_km2']),
        textcoords="offset points",
        xytext=(5, 5),
        ha='left',
        fontsize=9
    )

plt.title(f'Fire Area vs. Distance to Nearest Station\n(Pearson r = {corr:.2f},␣
 ↪p = {p_value:.3f})')
plt.xlabel('Distance to Nearest Fire Station (km)')
plt.ylabel('Fire Area (km²)')
plt.grid(True)
# Optional: Label top 5 largest fires
highlight_fires = analysis_df.sort_values(by='area_km2', ascending=False).
 ↪head(5)

for _, row in highlight_fires.iterrows():
    ax.annotate(row['FIRE_NAME'],
                xy=(row['nearest_dist_km'], row['area_km2']),
                xytext=(5, 5),
                textcoords='offset points',
                fontsize=9,
                weight='bold')
plt.tight_layout()  #  Prevents layout overflow
```

```
plt.show()
```



1.1 Conclusion

This notebook demonstrates a moderate positive correlation between distance to fire stations and fire area burned. Results may inform planning for new fire station locations or early detection systems in remote zones.