Wildfire Vulnerability Analysis Training Manual

Prepared For: Paul Keane

Project Context: Wildfire risk identification and analysis using geospatial and statistical techniques in Python and ArcGIS Pro

---

Overview

This manual documents the end-to-end process for preparing, executing, debugging, and exporting a geospatial wildfire analysis Python notebook that you developed and ran both locally and on an AWS EC2 virtual machine. It includes your technical challenges, how they were resolved, and recommendations for future workflows.

Environment Setup and File Management

Local Setup:

- Installed Anaconda Navigator.
- Opened `.ipynb` notebook directly **from the folder containing shapefiles and data**.
- Verified that Anaconda Navigator or Jupyter Notebook **does not need to be opened first**. Simply right-click and choose *Open With Jupyter Notebook* to preserve paths.

Initial Mistake:

- Opening the notebook from a different directory (or through Anaconda Navigator first) **broke relative file paths** like `California_Fire_Perimeters_2017.shp`, causing file-not-found errors.

Recommendation:

- Always unzip and run your notebook from the same directory where all shapefiles, rasters, and CSVs are located.

AWS EC2-Specific Setup:

- Made sure that the extracted ZIP project folder was fully self-contained on the remote desktop.

- Ensured Anaconda and Jupyter Notebook were launched using a path consistent with the project folder.

- Verified that all shapefiles and raster layers were accessible by Python from within EC2 without broken paths.

Path Fixes and Dataset Integration

Data Files Required:

- `California_Fire_Perimeters_2017.shp`

- `firestations.shp`

- LiDAR-derived rasters: DSM, DEM, Canopy Height (nDSM)

- 2020 StoryMap and prior analysis references

Fixes Made:

- Updated `gpd.read_file()` and raster paths to relative paths.

- Moved the notebook into the folder containing the data the key fix that prevented FileNotFound errors.

Geometry, CRS & Projection Challenges

CRS Errors:

- Encountered `ValueError` while setting CRS that already existed.
- Some geometries were invalid (`fire_centroids.is_valid.value_counts()` `False` entries).

Fixes:

- Used `GeoSeries.set_crs(..., allow_override=True)` only when necessary.
- Ran `fire_centroids = fire_centroids[fire_centroids.is_valid]` to clean invalid geometries.
- Reprojected everything to **UTM Zone 11N (EPSG:32611)** before calculating distances.

```python
fire_centroids_proj = fire_centroids.to_crs('EPSG:32611')
fire_stations_proj = fire_stations.to_crs('EPSG:32611')
```

Distance Analysis

Original Error:

- Incorrect use of `* 111` for calculating distances in kilometers from lat/lon (WGS84) without projection.
- Resulted in impossible values like `1.797693e+305` (float overflow).

Fix:

- Corrected method using UTM projection and `.distance()` in meters:

```python
fire_centroids_proj['nearest_dist_km'] = fire_centroids_proj.geometry.apply(
    lambda pt: fire_stations_proj.distance(pt).min() / 1000
)
```

- Verified correct distance range (~0180 km), resolved `NaN` values with cleaned geometries and valid projections.

Scatterplot & Correlation Plotting Issues

Initial Problems:

- Scatterplot failed to render due to PIL/matplotlib issue: `SystemError: tile cannot extend outside image`.

Fixes:

- Added margin padding to avoid rendering overflow.
- Cleaned data before Pearson correlation.
- Retained labels with adjusted offset:

```python
for _, row in highlight_fires.iterrows():
```

```
    ax.text(row['nearest_dist_km'] + 0.2, row['area_km2'], row['FIRE_NAME'], fontsize=9)
```

Pearson Correlation Result:

- Successfully calculated Pearson r and p-values between fire area and distance to nearest station.

- Plotted scatter with `sns.scatterplot()` colored by `YEAR_`.

Saving & Exporting Results

Save Strategy:

- Used **File Save and Checkpoint** to preserve notebook and outputs.

- Exported notebook as PDF via Jupyter: File Export Notebook As PDF.

- Alternative: used browser print to PDF (better format on some systems).

GitHub Push Instructions:

1. Export notebook as `.pdf` and save `.ipynb`.

2. Push both to GitHub.

3. Add `README.md` explaining structure, how to unzip, and run.

Final Interpretation Summary

- Fires with larger area tend to occur farther from fire stations but the correlation is weak.

- Nearest station analysis shows potential fire station service gaps.

- South/Southwest-facing slopes, steep slopes (>40%), and high canopy vegetation identify high-risk WUI zones.
- LiDAR DSMDEM subtraction provided Canopy Height model, enhancing the terrain-based risk mapping.

Full Challenge-to-Solution Timeline

Challenge	Fix/Resolution
Notebook paths broke	Launched notebook from correct folder
Shapefiles not loading	Corrected file paths
Invalid geometries	Filtered invalid with `.is_valid`
CRS errors	Used `set_crs(..., allow_override=True)`
Distance calculation overflow	Reprojected to UTM and used `.distance()` properly
PIL rendering crash	Adjusted matplotlib text labels
Pearson r was NaN	Cleaned distance data and recalculated
PDF text cut off	Used browser Print > PDF instead

Recommendations for Future Use

1. **Always run notebooks from the correct folder.**
2. Validate geometries immediately after `read_file()`.
3. Confirm CRS and projection before any spatial math.

4. Run full notebook before exporting to PDF.

5. Archive zipped version of all files used.

6. Push `.ipynb`, `.pdf`, and a helpful `README.md` to GitHub.

Status Summary

Task	Completed?
Local run with correct paths	
EC2-compatible ZIP package	
Geometry and CRS fixes	
Nearest station distance analysis	
Correlation and labeling	
Manual export options	
Personalized training manual	
Full retrospective challenge analysis	