SID's: 490612499, 480432823

Fire Risk Statistical Area Analysis in Greater Sydney

Dataset Description

Dataset Sourcing

Our team's final tables consisted of sections from a selection of databases including "Australian statistical areas (sa2) shape data, "suburb business and residential building details" and "NSW vegetation fire risk and categories", all sourced from course-provided, census-based datasets. Towards the end of the project period, a smaller, revised version of the "bushfire prone land dataset" was inserted in lieu of the original. NSW Elevation and Depth data, as well as a finer geographical statistical area dataset following "sa1" conventions were sourced as additions for further study(ABS, 2016) . These additions were useful in determining a new fire risk score for some areas to increase result granularity.

Given topography is amongst the three major factors influencing fire behaviour alongside fuel and weather (Department for Environment and Water, 2020), these data sets came in handy. "NSW Elevation and Depth data" included data regarding surface topography, including spot heights, relative heights, and contours (NSW Government Spatial Surfaces, 2019). However, fuel and weather data, as evidently major factors influencing fire behaviour, were considered and rejected as additional datasets on the basis of fuel data already being provided by the university, and weather data (e.g. rainfall, windspeed) being unactionable. The latter due to historical weather data being inconclusive as a forward-looking measure beyond long term climate trends, and real-time weather data leaving too little latency for action (i.e. by the time weather changes are shown, the weather occurrence is already in effect).Topography data however, is unlikely to experience major fluctuation year-to-year and is thus suitable to provide more immediately actionable information. Topography data was sourced from the NSW Government's Spatial Collaboration Portal (2019), first filtering by desired map layers and polygon boundaries, before download in JSON format with geospatial data in SRID 4326.

Pre-processing

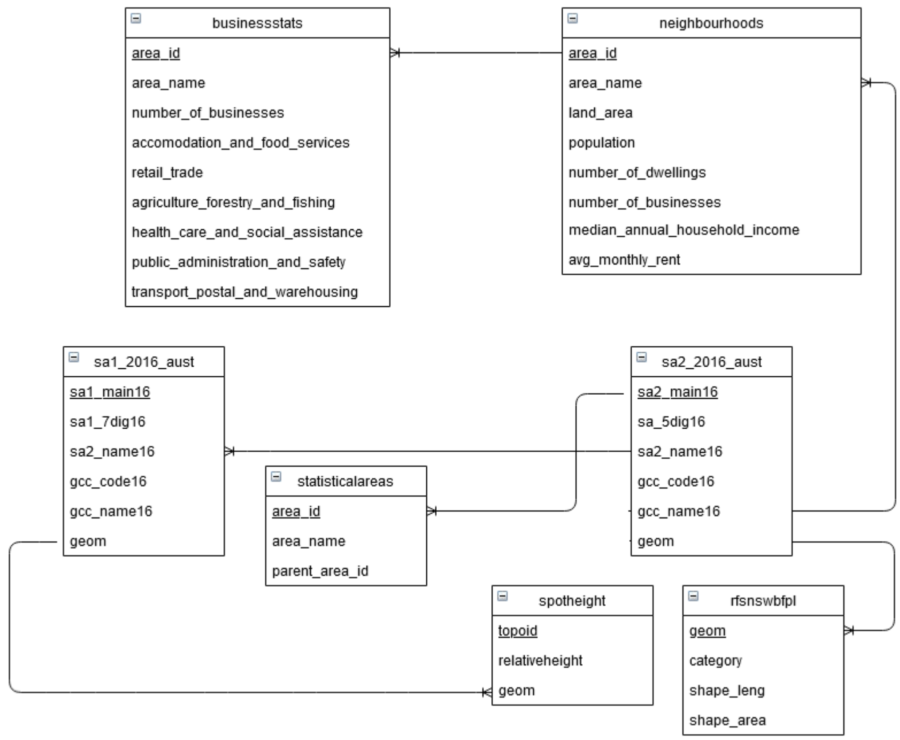
Data pre-processing for all datasets began with reading each file into Python through pandas functionality, converting to geopandas when applicable. Variable names at this initial stage were selected as future table names to avoid confusion, with "statisticalareas", "neighbourhoods", "businessstats", "rfsnsw_bfpl", and "sa2_2016_aust" each reading from correspondingly labelled CSVs and shapefiles. For the additional topography dataset, variable name spotheight was similarly selected to avoid confusion, and maintained throughout the remainder of the project, as that was the most pertinent attribute from a data analysis perspective.

For each dataset NULL values were then removed, and in the "sa2_2016_aust" dataset, all columns not pertaining to statistical area 2 were then dropped, excepting columns on suburb region (i.e. "gcc_name" and "gcc_code"), and the geometry. As geometry data in the original "sa2_2016_aust" shapefile was not consistently multipolygon data, and followed reference system SRID 4283, a replacement geom column was also created converting all values to multipolygons in SRID 4326 to maintain data consistency. The same process was followed with bushfire prone land data, similarly originally in SRID 4283 point data, as well as the topography dataset, to maintain data integrity. From the topography dataset all additional columns beyond "topoid" as the feature identifier, "geom" as the geometry, and "spotheight" as a relative height were also removed as they were irrelevant and to reduce load times.

For columns with data values to be used in further mathematical calculations and functions (e.g. population, dwelling, and business number, from neighbourhoods) writing conventions were corrected to maintain consistency (e.g. removing commas from ‘1,000’) before it is possible to load as a numeric data type.

Database Description

<u>1:1 Entity Relationship Diagram</u>



Regarding indexing, our team created a generalised search tree (GiST) index on the statistical area dataset (i.e. sa2_2016_aust[geom]), as in Appendix B. This was given that the sa2_2016_aust table was the one most frequently used in queries regarding bush fire prone land and topography, both containing extremely large quantities of data values. A GiST index was selected over GIN as the former is faster to build and update, with the inherent downside of potentially producing false positives eliminated by PostgreSQL automatically checking table rows to prevent this, as per PostgreSQL (2021) documentation. A second nonclustered index was created on suburb names (i.e. sa2_2016_aust.sa2_name16), to reduce execution time on queries without spatial data.

Fire Risk Score Analysis

Regarding risk analysis, calculation of fire risk scores largely followed the provided formula, albeit including integration of our team’s additional topography dataset, and a more granular sa1 map. However, as specified in the NSW Government’s Spatial Collaboration Portal (2019), the publicly available topography map *is* limited to 5m resolution (i.e. 5m per pixel), with “steep slopes, shadow and vegetation”, prevalent largely in urban suburbs, all interfering with LiDAR data collection. As a result, our team has separated the presentation of final fire risk results into a side-by-side overall Greater Sydney map, and a second, magnified map demonstrating higher granularity data, emphasising topography and bush fire prone land per suburb, as in Figure 1.

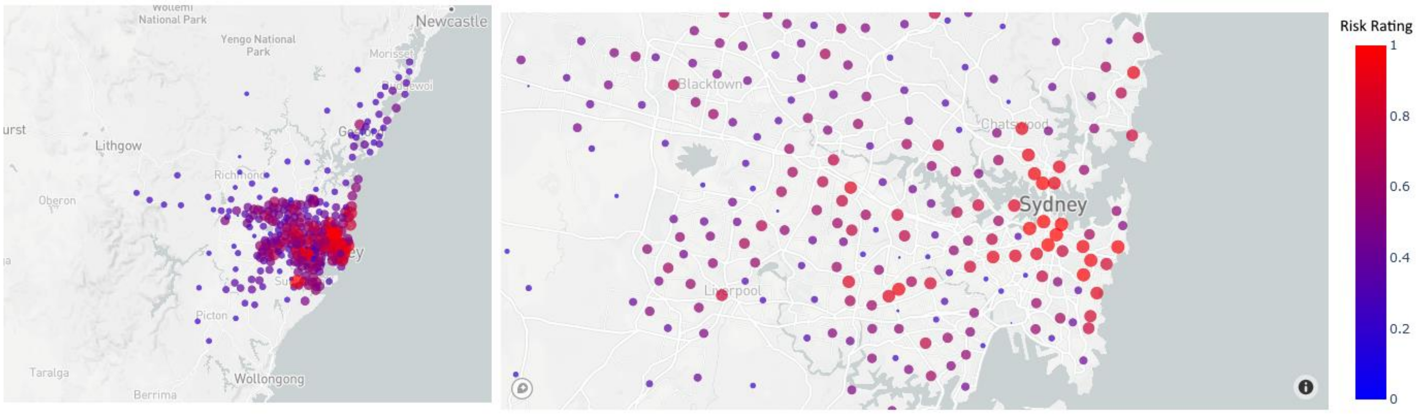


Figure 1: Side-by-side map representation of fire risk scores in Greater Sydney unmagnified and magnified to demonstrate higher granularity data on topography and bush fire prone land.

Regarding fire risk score calculations, SQL queries were employed to derive formula components for population, dwelling, business, and assistive service density, with assistive services further being defined as health and social assistance services available within each suburb. SQL queries were originally used to further derive averages, standard deviations, and Z-scores for each risk component, however due to query comprehension and legibility decreasing following the growing complexity of each subsequent query's (as seen in the nested queries in Appendix A), our team substituted this with Python's pandas data analysis library.

By presenting two versions of the fire risk map, our team attempts to account for a perceived fire risk disparity in more urban suburbs - namely that whilst more heavily built-up areas have been proven to decrease the risk of fire spread, this is moderated by the simultaneous effect of an increase in ignition chance (Price & Bradstock, 2014). In effect, the original formula skews fire risk towards factors of urbanisation by virtue of the sigmoid function incorporating four urban variables (i.e. Z-scores for population, business, dwelling, and assistive service density), compared to only a single environmental factor (i.e. bush fire prone land), whilst maintaining an equal weighting across all factors. This is demonstrated in Figure 1, where suburbs closer to the Sydney CBD demonstrate higher risk.

However, given rural areas have a much lower population density, and since Statistical Area boundaries are decided based on population, our team concluded that we should refine the results as depicted in Figure 1 with an additional measure to evaluate a more specific fire risk score for these areas based on finer sa1 data. From observing the data, we found records from the city were drastically different and swayed the rural z-scores, as represented in Figure 2. Our team was unable to incorporate this into risk scoring as formulaic weightings, however, as extant literature is unable to objectively define the degree of ignition and fire spread impact, merely to demonstrate respectively positive and negative correlations to increased urbanisation.

Fire risk as represented in the current Greater Sydney map supports previous discussion of high sigmoid function weighting towards more urban suburbs, with the Sydney CBD and surrounding suburbs recording the highest average risk across greater Sydney. Despite not being able to incorporate this discussion graphically, our team nonetheless recognises, following more detailed analysis incorporating sa1 boundary data, that an accurate fire score calculation *should* create a smoother distribution of fire risk - as would be expected by risk of fire spread increasing in rural suburbs, if not ignition chance. This also aligns with recent 2019-2020 bushfires, with a majority of the wild fires in, or around the Greater Sydney region occurring to the northwest: the Blue Mountains area.

Regarding the inclusion of separate rural Sydney fire risk results in higher granularity, from comparing fire risk between sa1 and sa2 boundaries, we found records from the city were drastically different and swayed the rural Z-scores, as represented in Figure 2. As literature and recent events have shown rural areas are no less prone to fire risk than closer to the city, as suggested by the provided sigmoid function, our team identified that recalculating the z-scores within the sample space of just these rural areas would give more accurate relative values for comparing neighbouring statistical areas within these regions. We attempted this through the addition of another positively weighted Z-score into the inner function, based on the relative topographical heights from these areas, however this was inconclusive, with risk similarly being skewed towards urban suburbs. Again, giving environmental factors a greater weighting provided more logical results, but any results we reached were ultimately arbitrary and unsupportable given previous discussion around the lack of research into appropriate environmental factor weightings. Consequently, we did not include these results in the overall fire risk.

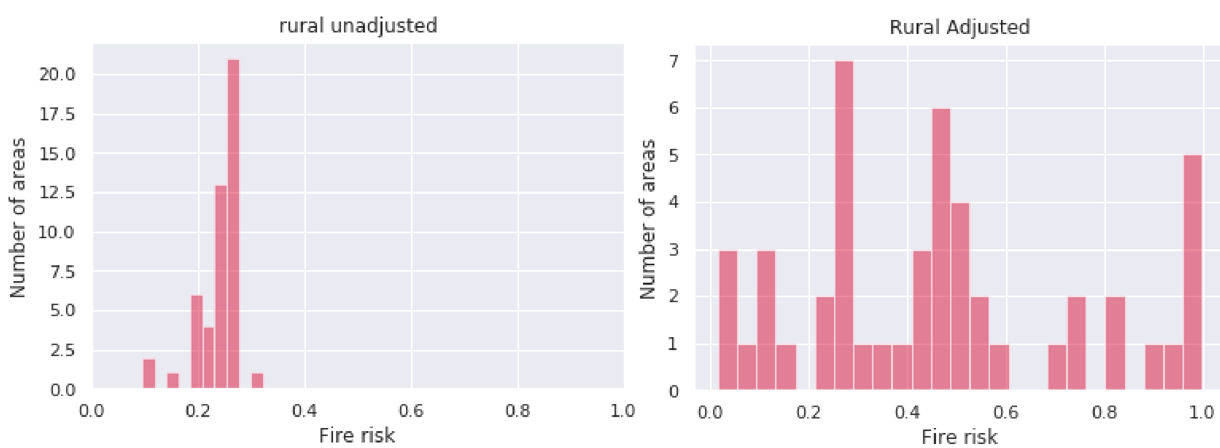


Figure 2: Histogram comparison of topographically and local-urban-density unadjusted and adjusted Fire Risk Scores

Correlation Analysis

Regarding fire risk correlation to suburb affluence, our team ran separate two-factor analyses comparing our final suburb fire risk scores to median yearly income and average monthly rent respectively, with results indicating fire risk is positively correlated with suburb rent, and negatively correlated to income. Correlation coefficients were derived by running separate calculations on our team's fire risk scores against income and rent figures from SQL queries. This was achieved through Python with Jupyter Notebook using an inbuilt pandas correlation function.

As in Table 1, initial results indicated fire risk increases with neighbourhood affluence, supporting the risk scores derived from a skewed sigmoid function but running counter to logic. Intuitively, this would suggest residents of wealthier suburbs are indifferent to and, in fact, seek suburbs with less fire safety. This prompted a second round of analysis separating the datasets by rent, noting a difference in correlation across the span of data.

Table 1: Initial correlation results using a Pearson Coefficient

Data - Fire Risk VS	Calculated Pearson Coefficient	Analysis
Median Income	0.199356	Minor positive correlation
Average Monthly Rent	0.3039458	Moderate positive correlation

Table 2: Correlation results following separation along high and low median income

Data - Fire Risk VS	Calculated Pearson Coefficient	Analysis
Average Median Income (Low Values)	-0.27987	Moderate negative correlation
Average Median Income(High Values	0.1190	Very Minor positive correlation

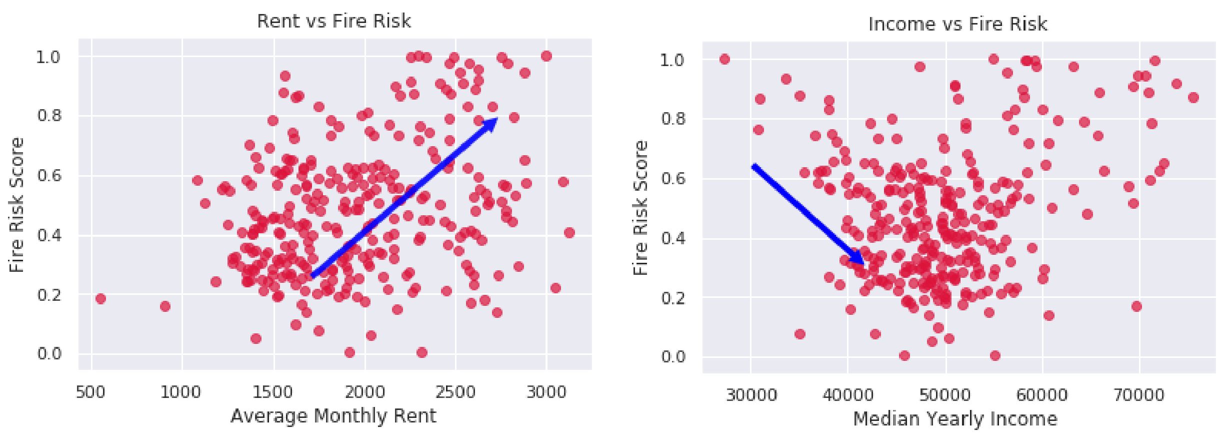


Figure 3;x, y laterally : Scatter plot comparison of average rent, median yearly Income, against the calculated fire scores for each SA2 suburb.

After correcting for correlation disparities between high and low median income suburbs (table 2), the final results indicated our calculated fire risk ratings were moderately correlated with suburb rent increase and inversely correlated with rent increase in low median Income areas (< $50,000 annually), whilst almost no correlation was seen in high median Income areas (>$50,000 annually).

Intuitively, this follows an explanation that lower Income areas have a higher risk of bushfire damage through less disposable income and zone tax to reduce fire risk and increased chances of fire spread (Price & Bradstock, 2014). However, wealthier suburbs also suffer from increased chances of ignition. Perhaps above a certain median income, high income earners can afford to live in areas with higher insurance premiums for fire damage. This is further supported by our monthly rent/fire risk correlation. Houses with higher insurance premiums would require higher rent in order for landlords to see adequate profit margins, although only a small percentage of australian houses are actually insured for fire damage (mozo, 2019) , this is enough to see a correlation between rent increase and fire risk increase.

References

Australian Bureau of Statistics. (2016). *Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2016*. Retrieved from: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202016

NSW Spatial Services. (2019). *NSW Elevation and Depth Theme.* Retrieved from: https://portal.spatial.nsw.gov.au/portal/home/item.html?id=f1b1b26b23ab4574a561f890edd3bd7e

Department for Environment and Water. (2020). *Fire behaviour*. Retrieved from: https://www.environment.sa.gov.au/topics/fire-management/fire-science/fire-behaviour

The PostgreSQL Global Development Group. (2021). 12.9. GiST and GIN Index Types. Retrieved from: https://www.postgresql.org/docs/9.1/textsearch-indexes.html

Price, O., Bradstock, R. (2014). Countervailing effects of urbanization and vegetation extent on fire frequency on the Wildland Urban Interface: Disentangling fuel and ignition effects. *Landscape and Urban Planning, 130*(1). 81-88. Doi: 10.1016/j.landurbplan.2014.06.013

O'Chee, K. (2019). Australian homes up to 66% underinsured, new research finds. Retrieved from: https://mozo.com.au/insurance/home-insurance/articles/australian-homes-up-to-66-underinsured-new-research-finds

Appendices

Appendix A - SQL Query for Z-score of Population Density

```sql
SELECT n.area_name, (((n.population/n.land_area) - (SELECT AVG(n.population/n.land_area) AS "avg_population_density"
                                                    FROM neighbourhoods n JOIN sa2_2016_aust sa2 ON(n.area_name = sa2.sa2_name16)
                                                    WHERE LOWER(sa2.gcc_name16) LIKE '%sydney%'
                                                    )
                    ) / ((SELECT SQRT(AVG((n.population/n.land_area - (SELECT AVG(n.population/n.land_area) AS "avg_population_density"
                                                                       FROM neighbourhoods n JOIN sa2_2016_aust sa2 ON(n.area_name = sa2.sa2_name16)
                                                                       WHERE LOWER(sa2.gcc_name16) LIKE '%sydney%'
                                                                        )
                                              )^2
                                             )
                                        )

                          FROM neighbourhoods n JOIN sa2_2016_aust sa2 ON(n.area_name = sa2.sa2_name16)
                          WHERE LOWER(sa2.gcc_name16) LIKE '%sydney%'
                          )
                         )
                    )AS "Z-score"
FROM neighbourhoods n JOIN sa2_2016_aust sa2 ON(n.area_name = sa2.sa2_name16)
WHERE LOWER(sa2.gcc_name16) LIKE '%sydney%'
```

Appendix B - SQL statement to create GiST index on sa2_2016_aust's geometry column

```sql
CREATE INDEX sa2GeomIdx
ON sa2_2016_aust
USING GIST (geom);
```