

Australia Fires

Group K

Project Members

Carissa Caccia (carissa.caccia@columbia.edu)

Anna Takacs (aet2161@columbia.edu)

Katie Aszklar (k.aszklar@columbia.edu)

Louis Nix (ln2419@columbia.edu)

Abstract

We are interested in looking at how the Australia fires developed over time, who was impacted during the fires and how news evolved around the topic within Australia and among other English speakers in the world. We will explore the following questions:

1. How has weather in Australia changed over time to support the thesis that climate change is contributing to increased and more severe fires?
2. What did the spread of fires look like during the 2019 fire season in Australia and who was impacted?
3. How did news spread in relation to the fires this past year both via news articles and twitter?
4. What was the general sentiment around the fires in 2019? How are topics and terms related in the discussion of the Australia fires and is there any relation between them?

We will make use of ggplot2, spatial techniques, text mining techniques and interactions in our visualization in a website format.

Data Description

Fire Data:

NASA satellite-collected point data with longitude-latitude coordinates and luminosity measures collected to measure fire occurrence. The data covers measurements of fires from from September 1st, 2019 to January 11th, 2020.

Census Data:

Demographics data from the Australian Bureau Statistics. From the data set we were able to get several levels of aggregation from granular dividing australia into over 2000 zones all the way up to population counts for the individual states. The data is in a shapefile format.

News Related Data:

Articles scraped from google, involving the scraped full texts from the articles and the publishing dates related to each.

Twitter Data:

Tweets using the query “bushfires” from June 2019 through February 2020 were scraped from Twitter, along with information on the user, hashtags, likes, and retweets. These were scraped using the python GetOldTweets3 library, which makes use of Twitter’s advanced search feature.

Weather Data:

Temperature, Precipitation, Humidity and Wind at 10m, 80m and 100m heights. This data will be available at the postal code level for Australia going back 10 years at a daily level. We are receiving this data from The Weather Source company.

Visualizations

Australia Climate Trends:

Our first visualization(s) will depict the climate trends in Australia for the past 10 years. This visualization will be used to “set the stage” for our spatial and text visualizations. With this visualization, we would like to convey the conditions present this year that led to the large amount of fires seen this year as opposed to previous years.

Spatial:

We aim to make two spatial graphs. The first is a point density plot (in the form of hexbins) indicating the fire intensity over time. This plot will show the intensity over time either through visualization or by use of a slider that a user can interact with.

The second spatial graph will be a choropleth map of demographic data in Australia. The goal with this plot is to show what populations may be most affected by the spread of the fires. As the demographic data contains multiple variables, this plot may also be interactive to allow the user to explore their own question.

Text over time (line):

We would like to create a line graph depicting two trends; the quantities of news articles covering the fires as well as the quantity of tweets about the fires over time. Additionally, this graph will display notable events such as the first fire report. It is our hope to gain insight into the reactions to the fires - was there an initial large spike with a huge drop off? Did the news spike, while tweets held their trend for longer, suggesting that though many have “moved on,” a significant portion of people continue to be affected by the fires?

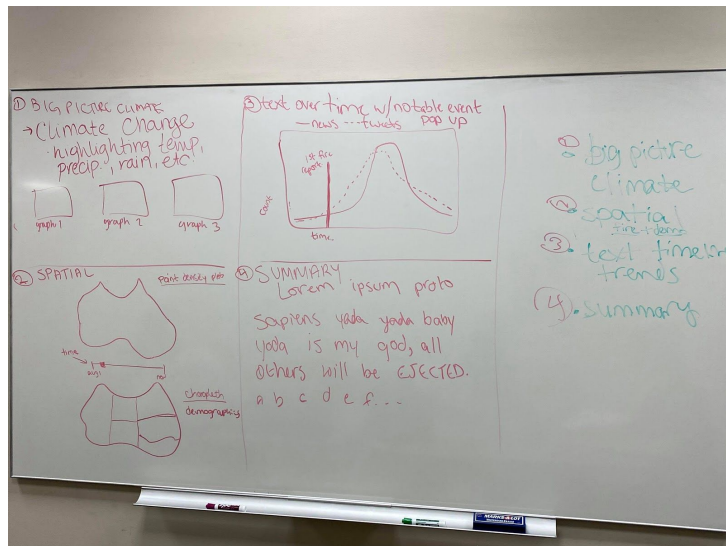
NLP Topics (line/network):

This graph would show the top 10 used words in the articles and their frequency over time. It would offer insight whether there was a shift in approach or discussion as fires have spread. For example, we expect to see that the word ‘donation’ has become frequent later than the word fire

but it peaked with a similarly large frequency. The network analysis would offer insight into how topics and terms are related in the discussion and whether there was any relation between, for example, the discussion of climate change and donations of wildlife.

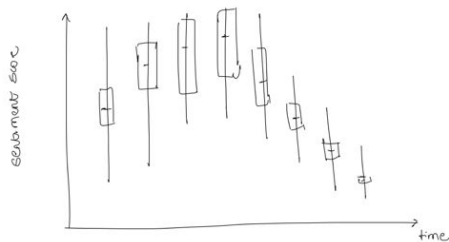
Brainstorming

Initial Overview:

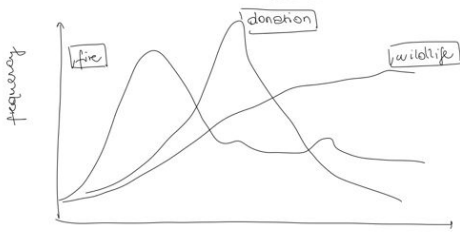


Additional Text Analysis:

1) how the sentiments of texts have changed over time



2) Topical analysis: top 10 most frequently used words



3) Network analysis: What topics are mentioned? Is there any specific focus?

