US Wildfire Data Visualisation and Analysis

Matt Bastiman, Jumaira Miller, Arran Taylor

December 2020

Abstract

Wildfires are an quickly increasing area for public safety concern worldwide and are involved in a feedback cycle with global warming. Fire suppression, prevention, and prediction are important tasks taken to address this problem. For countries as large and varied as the USA we seek to provide a visual tool for representing the enormous corpus of data regarding wildfires across the country in an attempt to aid wildfire response tasks. This proposed tool uses multiple coordinated views to present a dataset of over 1.8 million values succinctly and usefully. This is achieved using stacked bar charts, box plots; as well as an uncommon, state of the art, bubble plot modified to present geographic values. This tools efficacy is assessed by performing representative discovery and analysis tasks.

Contents

1 Introduction **3**
1.1 Motivation . . . 3
1.2 Goals & Aims . . . 4

2 Related Work **5**

3 Design Prototypes **8**
3.1 Alternative Design 1 . . . 8
3.2 Alternative Design 2 . . . 9

4 Proposed Solution **12**

5 Evaluation **15**
5.1 Frequency and Safety trends discovery . . . 15
5.2 Fire cause comparison . . . 16
5.3 Modelling and Presentation . . . 18

6 Future Work **19**
6.1 Animation Interpolation . . . 19
6.2 Improved Circle Packing and Placement . . . 20
6.3 Modelling Integration . . . 20

1 Introduction

1.1 Motivation

The devastating ecological, socio-economical, and atmospheric effects of wildfires are indisputable. Ecological concerns include the possible extinction of many endangered species and permanent changes to the landscape. For example, the population of Washington state's endangered pygmy rabbits are estimated to have been halved as a result of this year's fires[1]. Although the monetary repercussion of forest fires to society is difficult to measure, the socio-economic effects can be seen in the number of civilian casualties, blocking of public transportation and telecommunication networks, and the destruction of private and industrial infrastructure[2]. In the United States of America, the impact of the 2020 fires is estimated to be a staggering $20 billion[3]. Most apparent of all are the atmospheric effects of wildfires. Blankets of smoke pollute the air with harmful emissions which has the detrimental effect of accelerating global warming. Although it is commonly known that increasing frequency of forest fires are catalysing the effects of climate change, many people are unaware of the reverse also being true. In recent years, more scientists have come to accept that climate change increases the prevalence and intensity of forest fires just as much as forest fires increase the effects of climate change. Rising temperatures result in drier soil and vegetation. Warmer temperatures are also the cause of delayed snowmelt and longer droughts, which is another cause for drier conditions in forests. In both cases, the effects of climate change are fuelling more frequent fires and longer wildfire seasons. Carbon emissions increase with more prevalent and intense wildfire seasons, which in turn further accelerates the global rise in temperature. This vicious feedback loop implies a natural increase in the frequency and intensity of wildfires; however, the sudden inclination this year was unprecedented[4][5]. As of August 2020, the global frequency of wildfires has already seen a 13% rise in comparison to the entirety of 2019[6]. As for the intensity: Australia, Brazil, The Arctic Circle, and many states in America have reported record fires this year. As such, forest fires are currently a cause for major concern in many countries. Focusing on America, around 5 million acres have burnt across ten states. Amongst these states are California, Oregon, and Washington: all of which have reported unprecedented records in their history. Most casualties from forest fires are rarely from the fires itself, but rather from the fumes and smoke that they produce. Particularly in the year 2020, with the nature of the COVID-19 pandemic that targets our respiratory systems, the dangers of forest fires are ever more prevalent. As a result, there is a growing sense of urgency in the American Emergency Services – the Nation's support centres for wildfire firefighting in particular. Even more so when the President at the time, Donald Trump, has faulted poor forest management for the damage[3]. This immediate urgency to address and mitigate the impact of forest fires is the motivation for this project. The urgency demands more accurate fire risk assessment in order for the U.S. Fire Administration to take more effective preventative measures for the future.

1.2 Goals & Aims

The goal of this project is to design and implement an effective data-driven visual analytics system to support American fire services in making intelligent preventative decisions to mitigate the impact of future wildfires. Such measures include allocating funding to states accordingly. The user will be required to perform the following tasks to make accurate risk assessments:

- Observe the overall historical data of forest fires in the U.S.A to identify which states have historically been low-risk or high-risk areas.
- Be able to look at current data and identify the current most hazardous states.
- Predict high fire-hazard states based on past and present geospatial data on wildfire occurrences.
- View historical, current, and predictive data of fire-risk to get details on statistical data and identify the fire-hazard status of specific states.

Based on the anticipated tasks, the aim of this project is to use publicly available geospatial data of wildfires in the U.S.A to create a predictive model and visualisation of fire-risk mapping. There are several machine learning methods of modelling fire-risk, including logistic regression and through the use of neural networks[7]. For this project, we will be using a linear regression model to create a predictive dataset which will be used to visualise potential areas of high probable fire risk. Fire risk will be assessed based on the weighted average of the fire size where each fire is weighted to their size. This weighting is performed to account for the exponential danger caused by significantly larger fires. This average is then given a rating based on the NWCG fire classification guide[8]. To create an effective visualisation, two design ideas have been considered, before resolving on an optimal solution to meet the aims of this project.

2 Related Work

To gain insight into how others successfully went about visualising geospatial data, we referred to numerous papers to evaluate and compare methods for these visualisations to choose an appropriate method with regards to the designs of our own visualisations. These papers highlighted upon different methods along with evaluation metrics for how effective they were in achieving what was intended. I will now discuss some of this related work and talk about the inspiration for our designs. A rich source of information we first came across was 'The State of the Art in Map-Like Visualization' published in 2020[9]. This paper demonstrates that the choice of technique in supporting a particular visualisation is complicated as techniques are scattered across many different domains and gave us some ideas for the types of visualisation that would be beneficial for our dataset of choice. The authors of this paper demonstrated that for visualisations such as tube maps (i.e., London Underground), the lines for the train routes were always straight and associated with angles for changing the travel direction at different points[10]. Obviously, the route taken by the train would not be in a purely straight line for the majority of the train's travel, but this point is highlighted upon and we learned that some visualisations of geospatial data are tweaked in regards such as this to make interpretability of the tube map more easily readable for the traveller. This was the first piece of insight we took from this paper into realising that in order to display our geospatial data, we would have to consider the user's experience with the visualisation to make for an effective piece of software.

In the same paper, we learned about area schematisation which was revolutionarily insightful into our final outcome. Where area schematisation was discussed in the paper, we learned that geographic information and visualisations could be changed in appearance through size and colour to convey different statistical information. Figures were shown to highlight upon their use in the form of shape-deformed US states, of which the method was used to increase clarity and readability of perceiving data and identifying individual states. Immediately after learning this, we knew that we would have to display our states on a map in a way that allows the relevant users to quickly identify higher risk states and areas in the US without hassle. The benefit of the shape deformation technique used in Figure 21 of the paper is that the transformation correctly adapts the area of a region into recognisable shapes, increasing clarity.

Field schematisation was discussed in the paper, which involves the representation of mainly geographic data such as maps, of which are distorted based on certain conditions. This distortion is done in a few ways, mainly being through methods such as continuous stretching techniques as well as density techniques. The former method involved the distortion of local space based on field data, whereas the latter involves the aggregation of representatives for local regions also based on field data. We opted to use the continuous stretching technique for a few of our proposed designs in order to display a distorted cartographic map to the user based on each US state's fire risk. The following figure shows an example of this type of visualisation where the context is entirely based on being a map of Europe showing the spatial variation of seismic hazard.

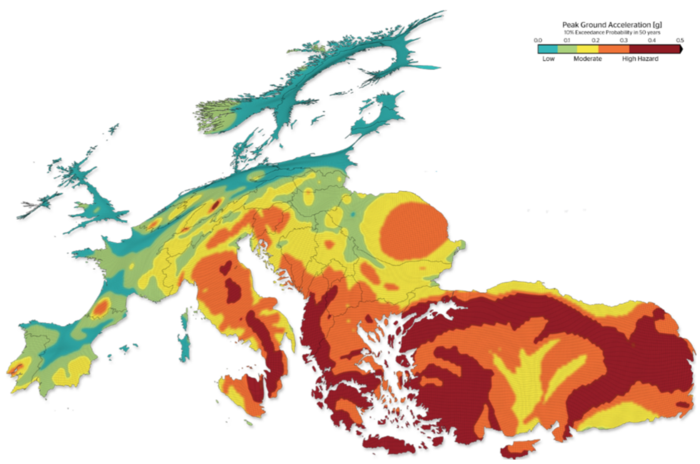

Figure 1: A continuous stretching technique distorting a map of Europe based on the spatial variation of seismic hazard.

Another piece of related work which inspired our project was 'A Machine Learning-Based Approach for Wildfire Susceptibility Mapping. The Case Study of the Liguria Region in Italy'[11] which highlights upon the benefits to using machine learning algorithms upon datasets such as the one we chose, in order to be able to predict data and create visualisations for risk perception of US states in years to come using the data we have up to the present day (or more appropriately, up to the most recent record in the dataset). Using these algorithms, we are able to think ahead and allow users to plan for preventative measures of wildfires far in advance.

The aforementioned publication looks at two main variables, having the x axis value plotted as the year (continuous data) against the frequency of wildfires for the corresponding year on the y axis. These variables of measure create suitability for a machine learning model, of which the dataset in the paper is trained in order to gain insight into trends and particular patterns within the data of which would not normally be easily interpretable. From this approach, we realised that we could easily apply a similar method to see possible future data, given the dataset being fit to a model. We ended up using the continuous attribute of 'FIRE_YEAR' in the dataset and plotting it against the 'FIRE_SIZE' variable in order to train the model to be suitable for prediction. Furthermore, we learned how we could test the validity of the model by using evaluation metrics and plotting the actual data against predicted data to visually be able to interpret how accurate the model was after learning of the importance of validity in the same paper.

It is important to note that for our final solution at tackling this aim we had; we did not end up visualising a typical cartographic map. We instead ended up using a packed circles diagram after learning of the superiority of this approach from the next paper: 'A circle packing algorithm[12]. This paper highlighted upon a method where data could be represented in a manner where (in our case) each state could be visualised as a circle, with two additional attributes being visualised based on the data itself: we took from this that we could have the size and colour of the circles representing two things, of which a combination of the two would allow the user to easily see which circles/states are most at risk based on these two characteristics.

The reason we used this circle-based visualisation is because not all users may be particularly familiar with the geography of the United States, which would create problems in identifying states quickly with something such as a cartographic map. Conversely, a packed circles diagram will much more easily highlight on only the necessary data where larger and darker circles will mean higher risk: interpretability and decision making of the potential users will drastically be improved with this in mind.

However, we wanted to take this approach to the next level: we saw the advantage of map-like visualisations in their ability to represent states with a degree of geographic accuracy, and we wanted to visualise the states somewhat according to their geographic location in combination with the packed circles diagram, and we learned of a powerful method which was able to do this: circle repulsion. 'Evolutionary Computation Solutions to the Circle Packing Problem' highlighted on this method[13]. This method involves originally placing overlaying circles according to their location, of which we encoded as their rough geographic location, and then repelling the circles away from each other so they only touch (similar to the typical packed circles algorithm) except with the additional element of being vaguely accurate geographically in their location in the visualisation.

These papers gave us great insight in tackling our goal, and we were finally able to start prototyping the designs and eventually come to a conclusion of a strong final design which we proposed and implemented for practical use as a powerful tool in preventing the spread of wildfire spread.

3 Design Prototypes

We proposed two design prototypes which I will now discuss in their entirety. For our designs, we had multiple sub-visualisations to show different views on the same data, with various forms of interactivity.

3.1 Alternative Design 1

One of our first designs proposed a way of seeing the spread of data with associated risk measures across the US in the form of a cartographic map, along with a box and whisker plot below it as shown:

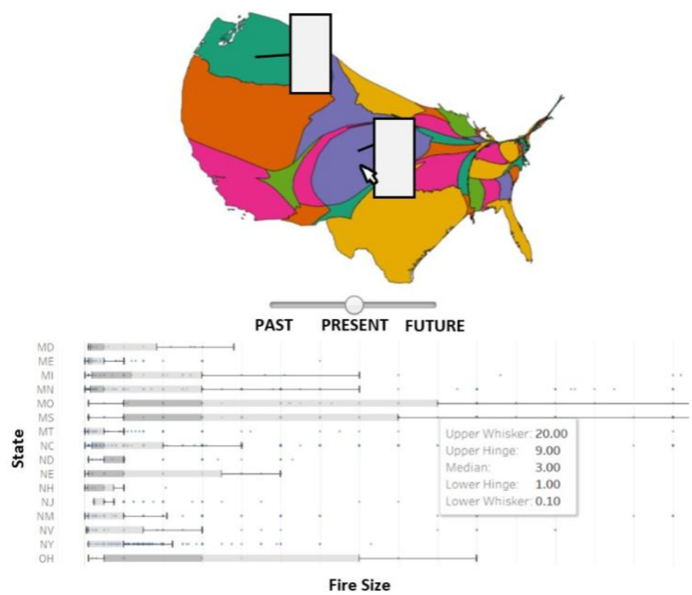

Figure 2: Design prototype involving the use of a cartographic map and box and whisker plot to convey important statistical information.

The distance cartogram is a powerful visualisation which changes the size of certain locations with respect to the data, as clearly shown. As we can see, western states are typically far larger and some states have almost disappeared entirely, displaying these states as being extremely low risk in comparison to the much larger states on the map. The user is also able to hover over different states to see more in-depth statistical information. When it comes to comparison, the user is additionally able to click on multiple different states which will retain the tool-tips on the visualisation to compare not only the size but different important statistical information to identify and eliminate anomalous data and draw further conclusions. As we can see, there is a slider underneath the map which allows the user to change the visualisation based on the decade of interest in the dataset: when a different decade is chosen, the idea is that the visualisation will change completely to represent data for the decade chosen. The 'future' data would be generated through a machine learning algorithm to give the relevant organisations the ability to prepare in advance for high-risk state wildfires, given that the model is of course appropriate and significantly accurate. The box and whisker plot underneath gives good insight into the data for the user to validate their decisions without uncertainty. Box and whisker plots are effective as they are easy to read and they can summarise the data from multiple states and display the results in a single visualisation, as opposed to generating a large amount of tool-tips when clicking on the states for comparison as mentioned before. The idea of this is to allow for easier and more effective decision-making. When certain state box and whisker plots are highlighted, they will show 5 important pieces of information: the upper whisker, the upper hinge, the median, lower hinge, and lower whisker.

3.2 Alternative Design 2

Our next design prototype does not include a cartographic map nor a box and whisker plot, but instead displays different visualisations altogether. This prototype was far more significant than the aforementioned prototype as we took more from this to implement into our final solution as I will now discuss in depth:

We chose to use coordinated multiple views in this prototype to allow the user to view different perspectives of the same data to draw more meaningful conclusions and make decisions with higher validity. As we can see, at the top of this prototype we have a packed circles diagram. This diagram shows a circle for each state, with the associated size and average fire-size class being the characteristics of the circles. For the colour, we chose to bin the colours dependent on the fire-size classes in a spectrum from yellow to red. Semantically, we learned that many people interpret the colour red

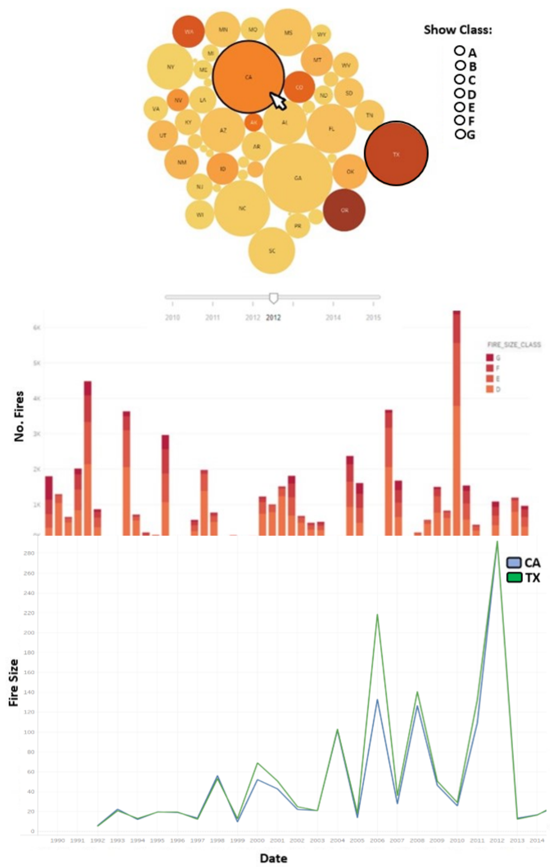

Figure 3: Design prototype involving the use of a packed circles diagram, a stacked bar chart and a line graph in coordinated multiple views.

with danger, and of course, fire and hence the darker the colour of the circle, the larger the associated average fire for that state. The size of the circle also represents the frequency of fires for that state. Radio buttons accompanied this diagram to filter out different classes of fires. Note that this aspect is very important to include, as many states would typically contain many minor fires, yet have very few if any significantly damaging large fires. This ability to filter the classes inherently allows the user to filter out anomalous data and compare the change of this filtering to the visualisation without filtering. This visualisation would change based on these filters that can be applied, while also changing based on the year chosen via the slider underneath. This slider would change the visualisation based on the 'FIRE_YEAR' attribute in the dataset, only including the associated year in computing the visualisation and of course, changing the visualisation as a result. The circles can additionally be clicked on to add another line to the line graph which we will soon discuss.

Underneath this diagram and slider, we have a stacked bar chart. This dissects the fire-size class information for each state to see the distribution and variance of fire classes for each state, with each class also being binned along the same spectrum of colour. The fire size classification is plotted against the frequency of these fires for the associated year for each state. As we can see in Figure 3, some states have far more G class fires than others, which is concerning. The user's ability to dissect the data in the packed circles diagram via this stacked bar chart creates more meaningful decision-making.

The line chart visualises continuous data on the x axis in the form of the date, against the fire size on the y axis. These lines only appear when circles are clicked on in the packed circles diagram to allow for comparison. Once the states are clicked on, the lines are visualised along with an associated legend at the top-right of the graph to show a key for the colour of the lines to avoid confusion. This creates a sense of historical insight into risk management to see if some states follow a consistent pattern, and if states are really worth investing into for preventative measures, while following the widely known details on demand mantra. These 3 components combined allow for powerful, easy, and quick decision-making while retaining necessary data.

4 Proposed Solution

The target user base is varied, as such the solution needs to cater to many arms and departments of the American emergency responders. It was important that the designs we used were able to give data broadly but could be filtered to the lowest useful level of granularity represented in our data.

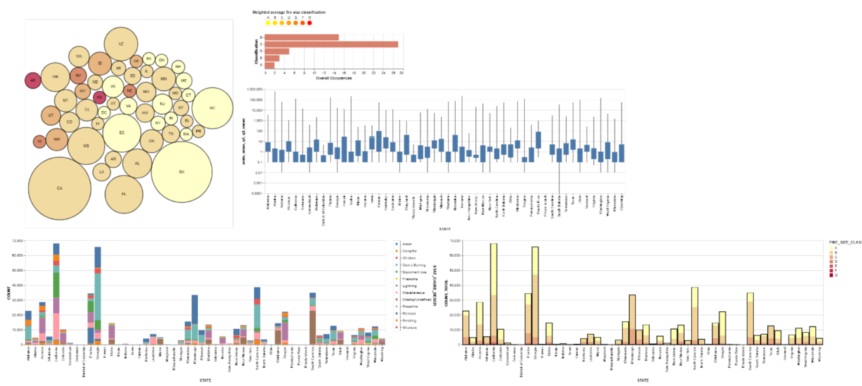

Figure 4: The final solution showing: the area schematized cartogram, fire category bar chart, box plot, and stacked bar charts.

Our data are geospatial in nature, as each fire is attributed a state, as well as a latitude and longitude for point of origin. The decision to display fire location aggregated by state rather than individual fire location was made because presenting a quickly interpretable summary of the large body of data was one of the primary goals of this visualisation. It was also determined that significant work would have to be undertaken to prevent occlusion as many fires occur in the same or similar areas, and the details of specific fires is not a goal this tool is stated to accomplish.

Our initial designs offered both a distance cartogram and a bubble plot. The distance cartogram has the benefits of abstractly representing the real location of the data. However it suffers greatly as distortion will make the original map largely unreadable, and this is compounded as a user must be familiar with the original map to understand to what degree any area is distorted. The cartogram also offers the advantage that more significant data are given more screen space, which makes investigation easier. This advantage may be a disadvantage when the user is, instead, interested in areas where fire risk is particularly low and may find it difficult to observe these areas. The bubble plot, by comparison, has a much greater ability to present comparative data by uniformly representing all values as circles.

The greatest flaw of the bubble plot is that it has no way of representing the geographic locations desired.

Our final compromise was the area schematized cartogram references by Hogräfer et al. [9]. This approach is a bubble chart where the bubble locations attempt to mimic locations in the data. In the figure provided we see California is in the bottom left and North Carolina is in the central right. This does of course represent limited geographical value however allows for enough general consistency that similarly positioned states appear nearby in the visualisation. This was important as we wanted to allow the user the opportunity to identify geographic trends in the data if present.

This chart contains two additional modalities: area, and colour. Area encodes fire frequency, which is the number of records for the given state (in the time period selected). Colour encodes for the fire danger of the state, with yellow showing least risk and red showing highest. These colours fit the "YlOrRd" palette suggested by ColorBrewer 2.0[14]. A spectrum with a red hue was selected as red is a common encoding for danger. These colours were also binned for consistency with the NWCG fire size classification.

Stacked bar charts are used to display data on fire causes and sizes by state. Vertical height is a simple and linear data encoding and bar charts provide an easy interface to read raw data as well as compare to other values. These were picked to allow a user to analyse specific values relevant to their investigation.

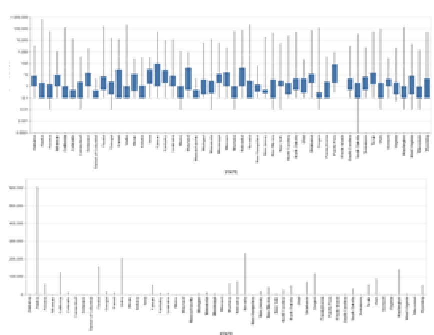

Figure 5: Box plot Y scale comparison with log_{10} on top and linear below.

Statistical data on fire size is presented as box plots. These are not novel visualisations, however are efficient in presenting relevant statistical ranges particularly for class based comparisons[15]. It was considered that box plots hold unrepresentative similarity when data is bimodal, however analysis of our data suggests that this is not a statistical reality in forest fires.

A log_{10} scale was used for the Y axis since experimentally a linear scale caused such severe truncation on the 1st to 3rd quartile that it became unreasonable to read any meaningful values from the plot. This is evident in the above figure and explained by the very few, very large fires which appear in the data effectively skewing the axis.

All charts were coordinated to allow for high precision data filtering, where the data could be viewed with regards to state and fire cause. These filters are applied by selecting the bar or circle relevant to the filter query, or the appropriate marks on the fire cause legend. The design pattern giving a broad overview of data then filtering was chosen to follow the Shneiderman mantra. Multiple coordinated views will also make it easier for a user to track the data they are interested in across different but related charts as only the highlighted data need be considered.

The final modality of the visualisation is animation and the ability to toggle data by time. This is achieved through radio buttons which allow the series to be transitioned through quickly, and for any frame to be isolated at will. All animation comes with significant drawbacks, these being the issue of 'eyes over memory'. It is typically better practice to allow a user the opportunity to observe all data they may be interested in rather than artificially occlude it through animated frames. The choice to include animation at all in this design was driven by the target goals, namely to identify trends in the data which may be beneficial for fire prevention, control, or suppression. It is less useful for the intended user to be able to see all data over the 23 year period represented than for them to be able to quickly see which trends (if any) appear when comparing historic to contemporary data.

The decision was made, however, to reduce the memory cost of the user by 'binning' the time intervals into 4 roughly equal time frames. This way macro trends could be tracked while smoothing the noise added by smaller inter-annual variations. This was also done in effort to reduce the 'Christmas light effect' where rapidly changing data becomes impossible to interpret because changes in marks occur too fast and with too little connection to each other. Frame interpolation was considered to ease this effect more, but was not implemented due to software limitations. Further work could benefit from pursuing this.

In an effort to aid the analytics of the data, we have used linear regression to include predictive data in the visualisation. This involved fitting a regression model to the data grouped by time period and by state (for a total of 108[1] models). These models are used to predict the frequency and fire risk for each state and a value for each was taken for the year 2021. This data is then used as a fifth time bin in the animation allowing the user to use the tool for displaying predictive data.

[1] Calculated as 52 * 4. Fifty-two states are counted as Washington DC and Puerto Rico are included in the dataset despite not being US States (at time of writing)

5 Evaluation

5.1 Frequency and Safety trends discovery

The tool provides a comprehensive means of data visualisation and exploration. Each section achieves or aids in a task which is relevant to the target user.

The cartogram, combined with the animation tool can be used to discover a five times increase in the number of fires in Texas (TX) between 2000-2015 compared to 1993-2000. This observation is made in comparing the circle sizes between the two years. In the first image Texas is one of the smaller circles, where in the second it has become the largest. The state's fire risk rating has not changed and so we can tell that while more fires have started in Texas they have not, overall, been of a greater risk than before.

Figure 6: Comparison of cartogram for data between 1993-2000 (left) and 2010-2015 (right) side-by-side

The cartogram would also seem to suggest an aggregation of "D" fire risk states in the north-west. In particular Oregon (OR), Wyoming (WY), Washington (WA), and Montana (MA). All of these states besides Wyoming have gone up a category compared to 1993-2000. This is corroborated when comparing the stacked bar charts for each state. With focus particularly on class "B" and above fires. While there may have been fewer overall fires, these now make up slightly more of the proportion of wildfires by each state.

For class "G" fires, those which affect an area of 5000 acres (20.2 km^2) Montana, Oregon, Washington, and Wyoming there was an increase from 31 to 59, 50 to 98, 27 to 61, and 10 to 35 occurrences respectively. These values can be obtained by over-over tooltips which give details upon demand.

From these observations, it can be concluded that the Pacific northwest of America is experiencing fewer more intense wildfires and so more funding may need to be allocated to suppression and large fire response than is presently available.

The proposed solution does present a few flaws while arriving at these conclusions, unfortunately. The first being the abstracted nature of state location. While the investigated states are all close to each other, they all border the state of Idaho, which does not appear connected to them in the 2010-2015 image. Despite this Idaho has been upgraded from a “D” to an “E” class and should be included in this observed trend. This trade-off of geographical abstraction to real location data could lead to trends being missed or data being overlooked. Future work could improve on the general accuracy of the area cartogram. Another clear issue is that Alaska (AK) which appears in red does appear to be connected to the states surveyed. It has been omitted because it is, in fact, not a part of the contiguous states, however prior knowledge of the map is required for this to be performed. The cartogram has clearly demonstrated its value above a standard bubble chart, however, and we feel justified in assuming some knowledge of US geography from our target users in US state, federal, or charitable departments.

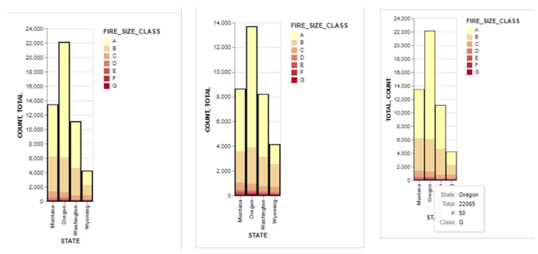

Figure 7: Fire size by state as a stacked bar chart comparison showing data between 1993-2000 (left), data between 2010-15 (center) and how a tooltip is used to extract exact figures from the comparison.

5.2 Fire cause comparison

The tool can be filtered by fire cause, making it a useful resource to find the state (or states) where certain causes are more prevalent and dangerous. For instance we see the cause bar chart filtered for fires caused by arson and fires caused by lighting.

Oklahoma (OK) leads the nation for fires caused by arson, where Oregon (OR) leads the nation in fires caused by lightning. Once filtered its possible to compare the two states. By interquartile range (visible on the box plot) Oklahoma’s wildfires are between 1 and 30 acres, where Oregon’s are between 0.1 and 0.3. Despite having generally smaller fires Oregon has far more than Oklahoma with 13659 compared to 9495.

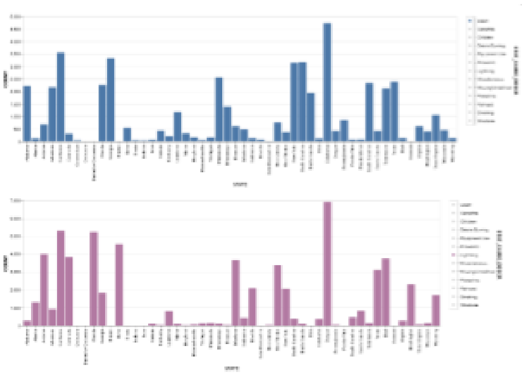

Figure 8: Fire cause by state for data between 2010-15 showing values filtered for arson (top) and lightning (bottom). Comparing bar height for each filter shows the prevalence of the cause in the fires for that state.

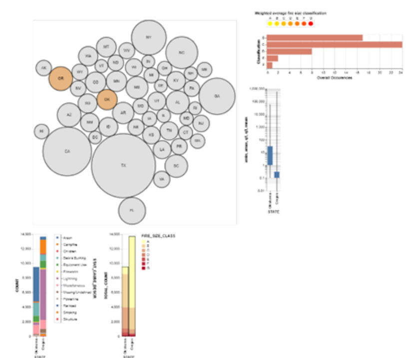

Figure 9: Data filtered for just the states Oregon and Oklahoma where fire causes can be compared in the bottom left, statistical ranges are taken from the bar plot to the right.

What can be understood is that, when accounting for all other causes of fire, arson was the cause of 49.8% of Oklahoma's wildfires in 2010-2015. Lightning was the cause of 50.8% of Oregon's.

This is valuable insight as the preventative measures which may be taken to reduce the frequency of fires in Oregon will not be as effective if applied to Oklahoma. There is also likely more call to reduce the rate of fires in Oklahoma as they are, when compared to Oregon, far larger and likely more dangerous as a result.

For this analysis it is sheer luck that Oregon and Oklahoma have relatively comparable areas (within the same order of magnitude for km^2). For most other causes it is simply states with the most fires overall which appear to have the highest incidence rates. The visualisation could have benefitted from normalisation with regards to state area, population, or fire frequency. The observations on Oklahoma and Oregon are demonstrative of the discoveries that can be made in the dataset and normalisation could allow for smaller states to show similar clear trends too. Without any normalisation it is unusual for smaller states to appear interesting in any aspect of that cartogram.

5.3 Modelling and Presentation

The predictive functionality of the tool through linear regression modelling both demonstrates how more sophisticated models could be used to display synthetic data alongside real data, as well as providing a simple predictive tool for data analysis.

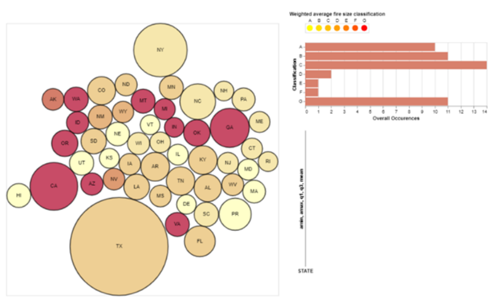

Figure 10: Cartogram for data predicted for the year 2021 by linear regression.

Initially it was considered that a single multidimensional model could be trained to represent the dataset as a whole, however this was quickly resolved to a solution where each synthetic datapoint would be generated by its own model. This was to avoid the curse of dimensionality as well as to provide accuracy metrics on a per-data-point basis.

Predictions were made for the state's fire frequency and the state's fire danger rating and a value for a future date (2021) was provided to obtain a predictive value. This means that states which grow in size compared to 2010-15 figures are on an upwards trend, where fires are becoming more frequent while shrinking states are on a downwards trend.

A similar observation can be made with state colour. A state getting redder would show that it is becoming more dangerous where a yellower state is becoming safer.

Because the user will already be familiar with the visualisation, the fitting of non-real predictive data allows for a non-technical user to analyse and interpret the output of a machine learning model as if it were any other data. The inclusion of synthesised data in our solution demonstrates that 'real' data can be combined with modelled data seamlessly for user interaction.

By the National Interagency Fire Center statistics for 2019[16], the models identified the top 4 states by cumulative wildfire size as rank "F" or "G".

The effectiveness of each model is also assessed by its R2 score. This score represents the (squared) distance from the linear regression line to the points of the model, where a lower score is a less accurate model. Since the average score was 0 it can be safely concluded that the models were not especially effective for predictive purposes.

This was, however, expected. Wildfire frequency and size are much too variable to be effectively represented by a linear function. This aspect of the visualisation simply served to represent the overall trend in the data which naturally could be approximated from viewing the previous animation frames.

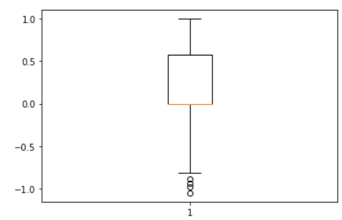

Figure 11: Box plot of R^2 score for each 108 models. Lower values indicate a "worst" fit for the data.

6 Future Work

The tool presented offers a great deal of overview, depth, and flexibility for considering the dataset. Given the potential that continued development offers to protect property and lives from fire, the following areas are offered as avenues for future work with this tool.

6.1 Animation Interpolation

The animation of the cartogram and charts is sudden and not easy to track. Effort has been made to attempt to keep the frames simple and limit unnecessary movement of marks by ensuring that filtering is preserved between frames and that states maintain a similar position in the chart. This could be improved significantly by implementing smooth frame transitions allowing the user to track the changes in size, colour, and position with more ease. The Christmas light effect would be reduced even further by also making time bucketing dynamic, so that the user can choose to bucket by their own time ranges depending on their area of interest. This chart is implemented in Altair and this does not seem the appropriate platform for continued development down this avenue, as there is little to no native support for animated transitions at the time of writing.

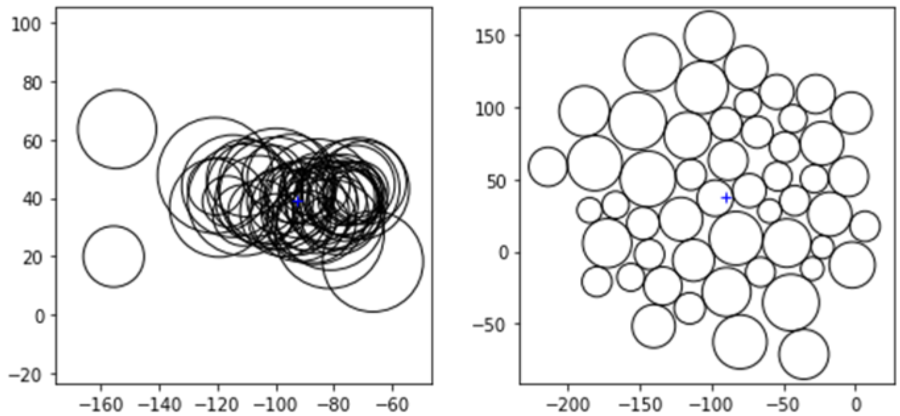

Figure 12: Demonstration of applied circle packing technique. Data is initially encoded allowing for overlap (left) and then repulsion and attraction is applied until all points are separate (right)

6.2 Improved Circle Packing and Placement

To achieve the desired area cartogram, a technique of placing circles, inducing iterative pairwise repulsion, and then collecting the points was implemented. This technique is crude, slow and messy, resulting in inefficient packing, inconsistent placement of nodes and intersection bugs. This technique should be improved upon to develop an efficient circle packing algorithm with variable circle size, where a node's location is honoured to the highest degree possible. One consideration is to build upon a graph-based system similar to the R implementation found here[17].

6.3 Modelling Integration

For this, and many more, time series visualisations - a generic approach to implementing data driven machine learning models will allow for higher quality user-end predictive analysis. In this tool we demonstrate the use of a simple model for trend following on a subset of the presented attributes. Wildfire modelling produces incredibly sophisticated simulations of the effects and conditions of wildfires. Including an intelligent model in a system such as this could provide a user far deeper insight into the predictive power of their data without needing to leave a tool they are comfortable using.

References

[1] E. Pennisi, "As wildfires continue in western United States, biologists fear for vulnerable species," 30 September 2020. [Online]. Available: https://www.sciencemag.org/news/2020/09/wildfires-continue-western-united-states-biologists-fear-vulnerable-species.

[2] C. A. J. M. Rábade, "Social Impact of Large-Scale Forest Fires," in Proceedings of the Second International Symposium on Fire Economics, Planning, and Policy: A Global View, 2008.

[3] BBC News, "California and Oregon 2020 wildfires in maps, graphics and images," British Broadcasting corporation, 17 September 2020. [Online]. Available: https://www.bbc.co.uk/news/world-us-canada-54180049. [Accessed 4 December 2020].

[4] T. M. K. L. N. Harris, "6 Graphics Explain the Climate Feedback Loop Fueling US Fires," World Resources Institute, 16 September 2020. [Online]. Available: https://www.wri.org/blog/2020/us-fires-climate-emissions. [Accessed 4 December 2020].

[5] Union of Concerned Scientists, "The Connection Between Climate Change and Wildfires," Union of Concerned Scientists, 11 March 2020. [Online]. Available: https://www.ucsusa.org/resources/climate-change-and-wildfires. [Accessed 4 December 2020].

[6] World WildLife, "2020 Forest Fires Globally Could be Worse Than 2019, WWF Warns," 27 August 2020. [Online]. Available: https://www.worldwildlife.org/press-releases/2020-forest-fires-globally-could-be-worse-than-2019-wwf-warns. [Accessed 4 December 2020].

[7] D. Stratoulias, R. Landenberger, N. Adaktylou, "Wildfire Risk Assessment Based on Geospatial Open Data: Application on Chios, Greece," International Journal of Geo-Information, vol. 9, no. 516, 2020.

[8] The National Wildfire Coordinating Group, https://www.nwcg.gov/data-standards/approved/fire-size-class 8 December 2020

[9] Hogräfer, M., Heitzler, M., & Schulz, H. J. (2020). The State of the Art in Map-Like Visualization. Computer Graphics Forum, 39(3), 647–674. https://doi.org/10.1111/cgf.14031

[10] Wolff, A. (2007). Drawing subway maps: A survey. Computer Science - Research and Development, 22(1), 23–44. https://doi.org/10.1007/s00450-007-0036-y

[11] Tonini, M., D'andrea, M., Biondi, G., Esposti, S. D., Trucchia, A., & Fiorucci, P. (2020). A machine learning-based approach for wildfire susceptibility mapping. The case study of the liguria region in italy. Geosciences (Switzerland), 10(3). https://doi.org/10.3390/geosciences10030105

[12] Collins, C. R., & Stephenson, K. (2003). A circle packing algorithm. Computational Geometry: Theory and Applications, 25(3), 233–256. https://doi.org/10.1016/S0925-7721(02)00099-8

[13] Flores, J. J., Martínez, J., & Calderón, F. (2016). Evolutionary computation solutions to the circle packing problem. Soft Computing, 20(4), 1521–1535. https://doi.org/10.1007/s00500-015-1603-y

[14] C. Brewer, M. Harrower, B. Sheesley, A. Woodruff and D. Heyman, "ColorBrewer 2.0," [Online]. Available: https://colorbrewer2.org/.

[15] D. Williamson, R. Parker and J. Kendrick, "The box plot: A simple visual method to interpret data," Annals of internal medicine, no. 110, pp. 916-21, 1989.

[16] National Interagency Fire Center, "Historical year-end fire statistics by state," [Online]. Available: https://www.predictiveservices.nifc.gov/intelligence/2019_statssumm/fires_acres19.pdf.

[17] M. Bedward, "Graph-based circle packing," 12 12 2020. [Online]. Available: https://cran.r-project.org/web/packages/packcircles/vignettes/graph_packing.html.