**COMP1730/6730 S2 2018 Project Assignment**

**Semester 2, 2018**

**Report**

**Question 0: Data and design**

1. The maps are stored as CSV files. This is not the only way that the data can be stored. Suggest another way, and explain what its advantages and disadvantages would be.

Ans. The other way to store maps is JSON data as it also supports large volume of data and supports hierarchical information.

Advantages of Jason over CSV

* Its lightweight, compact and versatile.
* Json is much faster while reading large amounts of data.
* It is better at showing **hierarchical / relational** data.
* CSV will lose data if we have to save data regarding a product. Eg. If we have to save many reviews about a product. We will be forced to make some cut-off for how many reviews we show . If we don’t then each row will have different number of columns which would make parsing impossible.
* It is easier to read JSON data than CSV because the default CSV reader application sometimes split to adjacent columns which makes it unreadable while **working** with huge amounts of data.
* Json is the default choice when working with data at scale as it supports modern API’s and RESTful services

Disadvantages

* If used improperly can pose a security hole (don't use eval), Not all languages have libraries to interpret it.
* If we have a simple schema with flat data then CSV would be a better choice as we will end up having a large file in JSON ,if size of data is of any concern.

1. In Question 1, we will load each of these files into a lists of lists. Describe some advantages and some disadvantages of using a list of lists to represent the data.

Ans. Advantages

Python lists are very efficient containers as they specialize in efficient insertion, appending, deletion and concatenation. They are easy to construct and manipulate because of their comprehensions.

Lists are numerically keyed and can be sorted and have values removed or added.

Disadvantages

When we try to handle a multidimensional list with a lot of data then it is slow.

Element wise arithmetic operations is a bit complex when compared to numpy.

Does not have an easy way to do arithmetic operations like in numpy

Getting the indexes of an element in the list is very difficult.

1. Suggest another data structure we could use to represent these files. Compare it to the list of lists. What are the advantages and disadvantages of your proposed data structure?

Ans. We can rather use Numpy

Advantages

Numpy is efficient and more convenient we can do vector operations and matrix operations very easily and efficiently.

And because of the inbuilt functions we can do convolution, fast searching , basic statistics etc

Disadvantages

Numpy array requires elements with same data type.

There are in built functions to read and write only for certain file formats.

1. We can use this data set and use fire risk function to map heat areas to get the places with more cycle accidents

https://www.data.act.gov.au/Transport/Cyclist-Crashes-Heat-Map/8ea8-wvzq

Question 1: Loading the datasets

1. What type does each of your file-loading functions return?

All the four function returns list of lists

load\_vegetation\_type returns list of strings

load\_vegetation\_ density and load\_wind\_speed both return floats but no value tues are appended with “None “

Load bushfire just returns a list of 0’s and 1’s and blanks

1. Explain how you accounted for blank values, and why you chose to account for them in the way you did.

In load vegetation type we converted to strings as the data which it had could be converted to that and also to make it easier to workwith them we chose to convert them to string.

In load vegetation density and load wind speed we converted to float to keep the values with utter precision as we did not want to loose precision by converting to string

In load bushfire we kept it as it is as we felt it was much easier to handle them that way rather than converting it to any other data type.

1. How many blank values are there in each file?

Question 2: The maximum wind speed

1. What is the highest wind speed in each map?

Ans. Highest wind speed

In anu is 7.29959

In south is 7.0812

In act is 8.60718

1. Since the list returned is list of lists with x number of lists and y number of elements in each list we get time complexity as O(x\*y)

Question 3: The most common vegetation

1. For anu

Open Woodland : 50

Woodland : 125

Grassland : 65

Forest : 370

Open Forest : 368

Urban Vegetation : 315

Golf Course : 0

Vineyard : 0

Pine Forest : 0

Shrubland : 0

Arboretum : 26

For south

Open Woodland : 10950000.0 sq m

Woodland : 15910000.0 sq m

Grassland : 519800.0 sq m

Forest : 440000.0 sq m

Open Forest : 2016000.0 sq m

Urban Vegetation : 0.0 sq m

Golf Course : 0.0 sq m

Vineyard : 0.0 sq m

Pine Forest : 0.0 sq m

Shrubland : 0.0 sq m

Arboretum : 0.0 sq m

For act

Open Woodland : 46154000.0 sq m

Woodland : 59670000.0 sq m

Grassland : 3458200.0 sq m

Forest : 37100000.0 sq m

Open Forest : 29856000.0 sq m

Urban Vegetation : 0.0 sq m

Golf Course : 0.0 sq m

Vineyard : 0.0 sq m

Pine Forest : 0.0 sq m

Shrubland : 0.0 sq m

Arboretum : 0.0 sq m

1. For anu

Open Woodland : 100000.0 sq m

Woodland : 625000.0 sq m

Grassland : 13000.0 sq m

Forest : 3700000.0 sq m

Open Forest : 2944000.0 sq m

Urban Vegetation : 0.0 sq m

Golf Course : 0.0 sq m

Vineyard : 0.0 sq m

Pine Forest : 0.0 sq m

Shrubland : 0.0 sq m

Arboretum : 0.0 sq m

For act

Open Woodland : 23077

Woodland : 11934

Grassland : 17291

Forest : 3710

Open Forest : 3732

Urban Vegetation : 4023

Golf Course : 712

Vineyard : 95

Pine Forest : 6728

Shrubland : 705

Arboretum : 260

For south

Open Woodland : 5475

Woodland : 3182

Grassland : 2599

Forest : 44

Open Forest : 252

Urban Vegetation : 1866

Golf Course : 336

Vineyard : 0

Pine Forest : 325

Shrubland : 30

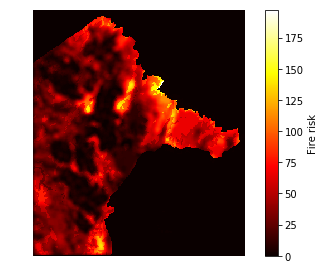
Arboretum : 43

1. We will index n\*n times.

Question 4: Fire risk

1. Describe the similarities and differences between your fire risk map and the BPA map. Explain how your calculation causes these similarities and differences.

Ans.



The similarities are the fire risk map clearly shows the northern-eastern part of map as shown in BPA. It clearly shows the patchy areas fire wasn’t prone like in areas of Molonglo Valley, Weston Creek and Tuggeranong where fire is not prone as shown in BPA with black colour which means its not prone . It also shows other areas with no names with pitch black which means they are not prone at all. The differences are southern areas of the map are not shown at all .Areas like Coree, Booth, Rendezvous Greek, Cotter River and Mount Clear are not shown at all. In our code we have taken the approximate value of the wind speed and get the details of the nearest cells into an array then we calculate the risk factor based on the type of vegetation . Hence areas with shurbs and pine have higher velocity of wind and more fuel to burn because of the type of vegetation hence they are mostly red and yellow as shown in BPA. And areas in black are calutated on the fact that there are no vegetation in those areas hence dry land with no vegetation has no fuel to fire hence they are black and fire risk is 0.

1. Also in your report, explain how you handled the calculation of fire risk at the edges of the map and why you chose to handle the edges in this way.

Ans. In our code if the vegetation density for a particular area is blank then we just get the fire risk as 0 as there is no vegetation in those areas. So we check if the vegetation density in the CSV has any value or not and if it doesn’t have any value then the value of fire risk is set to 0 hence this results in black colour at the edges of the map.

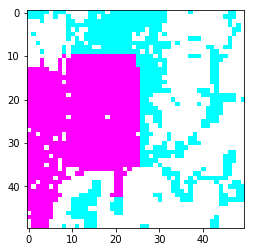
Question 5: Simulating the spread of bushfire

1. Write down how many steps you simulated

Ans. 26

1. Include an image of the resulting bushfire (remember, you can use show\_bushfire to make an image).

Ans.



1. Compare the spread of fire produced by your simulation qualitatively to the provided map of the real 2003 bushfire. By comparing qualitatively, we mean that you should compare the two plots and summarise in your own words in what ways they are similar and in what ways they are different.

Ans. By comparing initial\_2003\_bushfire and 2003\_bushfire we are getting the same plots so hence our simulation works same with respect to the what BPA ACT .

Question 6: Comparing to the 2003 bushfire

1. Write down the accuracy, as measured by your implementation of compare\_bushfires, of your simulation result from Question 5.

0.4353448275862069

b. How good would you say your simulation is?

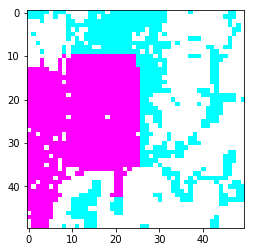
I feel our simulation is mostly similar to the real bush fire as we got similar maps when we ran on both the anu maps

Question 7: Simulation with fire risk

1. Write down how many steps you simulated.

Ans 25 for anu map

1. . Show an image of the resulting bushfire (remember, you can use show\_bushfire to make an image).



c. Compare the resulting map of fire affected areas to the real 2003 bushfire both qualitatively (by eye) and quantitatively (using compare\_bushfires).

By eye the simulations looks similar to the real bush fire and by using compare\_bushfire we got 0.4336618650492798 which is almost similar to the original value of 0.435(south).

d.Is your simulation realistic?

Yes since it is almost similar to the original simulation it should be realistic.