Singapore's Historical Climate Analysis

IS428 Visual Analytics and Applications

Written by Climate Vizards

Wendy Ng Sock Ling

Harvey Kristanto Lauw

Sin Myeong Eun

Abstract—Anyone who has visited or lived in Singapore would wonder, what is the point of analysing Singapore's climate when its land size is too small for its citizens to ponder over the differences in its regional climate? It is indeed true that the country's weather differences in different areas are relatively trivial, as compared to bigger countries where drastic differences can be observed. However, it would be relevant and useful for the residents in Singapore to know how their country of residence is experiencing any climate change over the past 10 years. Another useability of this research would be for those property buyers who definitely would want to look for housing at cooler and less rainy region. There were other trends that were spotted using various visualisation techniques as well, which will be elaborated in the subsequent sections.

1 Problem

Meteorological Service Singapore's website, weather.gov.sg, has a huge volume of climate data but the website does not make use of it in the most insightful way possible and as such,viewers would find it difficult to understand the visualisations at the first sight. The data can be better utilised if the visualisation techniques used were to be improved.

We, the Climate Vizards, aim to improve the visualisations on the website so that Singapore's historical weather data can be viewed at greater ease for those who might be concerned with it, such as people who are looking into purchasing a flat at slightly cooler and less rainy regions.

2 Motivation & Objectives

In the national day rally 2019, PM Lee mentioned that "Climate change may seem abstract and distant for many of us, but it is one of the gravest challenges facing humankind". [1] Moreover, the threat of rising sea level due to factors such as rising temperature was also mentioned. His speech regarding the issue of rising sea level caught our attention and we wished to speculate whether the statistics given were indeed true. However, we were unable to retrieve relevant sea level data. Hence, we have decided to look into how climate such as rainfall and temperature have changed over the years in Singapore, because despite its minute land size, even a single change in temperature or rainfall can be noticed by the locals. While scanning through weather.gov.sg, we noticed there were some visualisations that we concluded to have utilised less appropriate plots. For example, the yearly visualisations of temperature did not provide much insightful information of the annual mean temperature of Singapore as it simply used Singapore's geological map to show Singapore's temperature as a whole. This is evident from Figure 1. Our aim is to present the set of data by using a more appropriate plot such as calendar heatmap to show an overview of annual mean temperature in Singapore and illustrate any pattern that can be viewed from the plot.

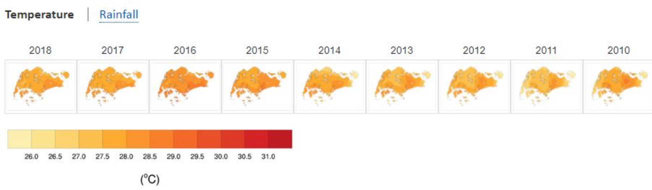

Figure 1

As for the visualisation that the website provided for monthly temperature/rainfall, it did not consider the fact that the data only represents the temperature/rainfall at the specific station, meaning that the visualisation contains uncertainties arising due to the lack of interpolated data points. This was one of the major problems that we have identified, because it'll provide inaccurate or misleading information to the users.
As such, we felt the need to explore some GIS packages such as spastat,gstat and raster. The use of these packages will be elaborated in our analysis of interpolation of the data.

3 Approach

3.1 Data Collection

We retrieved the necessary temperature and rainfall data from Meteorological Service Singapore's website [2]. The range of time period of the retrieved monthly data is from November 2009 to October 2019.

3.2 Data Preparation

3.2.1 General Preparation

We leveraged on lubridate package to do some data preparations. Firstly, we created a date field using the fields Year, Month and Day given. Afterwhich, a month field with the month's abbreviation was created.

As for the temperature data, we replaced all the minimum temperature of 0 degree celsius with N/A values as it is not possible for temperature to be 0 degree celsius in sunny Singapore. We then merged the temperature and rainfall data with station records which comprise of station coordinates with the use of left join. Lastly, we replaced all the whitespace in all fields names with underscores.

3.2.2 Calendar Heatmap

Based on the date field that was created in the general preparation steps, we found the day number. of the week in date and store it as week day. We had to convert this variable to a weekday factor . Afterwhich, we found the week of the year for each date and normalised the values to represent the week number of the specific month of the date by using week.of.month() function inside a for-loop.

For temperature heatmap, we skipped the N/A values in temperature data by using na.omit() function. As for rainfall, there was no N/A value so we did not need to deal with it.

3.2.3 Geospatial Interpolation

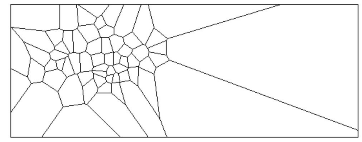

Figure 2

We noticed that none of the Climate data retrieved from weather.gov.sg comprised of coordinates required for geospatial study, hence a left join with the stations' coordinates on a separate CSV file was required. For the purpose of this study, we will not be leveraging on predefined spatial polygons which can be sourced online, due to the unique coordinates that were given in the station records.. We will instead use the thiessen polygons which can be created using Spatstat package as shown in Figure 2. [3]

Figure 3

Before we proceed with the thiessen polygon, we need to make an intersection with Singapore's boundary polygon in order to mark out the areas of interest for interpolation. Before that we had to confirm the alignment of bbox and proj4string of both objects, which will produce an output shown in Figure 3 above.

We then moved on to create empty grids with the existing list of coordinates from the joined data file. Our group chose to replace any NA values with median value as Gstat package to be used for interpolation requires valid data. Only then we can perform Inverse Distance Weighted Interpolation (IDW) to fill every empty grid with the variable of interest at every station location. Fine tuning in this analysis will not be done as most of the data will be replaced with median values and the range of data is not significantly big, hence it'll be assumed to have minimal error.

Once IDW is completed, the interpolated data is fitted into the 1st order polynomial in order to be geo-referenced. Although the trend might diminish some movements of the image, it will provide us a more consistent image to work with.

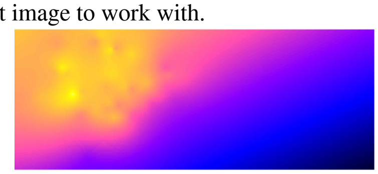

Figure 4

Lastly, kriging is performed to reduce the residuals in the data points by computing the 1st order polynomial model into the variogram model which will complete the interpolation of data. This helps to generate a kriged surface which will take into consideration of the station locations and fit accordingly to the trend as shown in Figure 4.

3.2.4 Monthly Distribution Across A Year

3.2.4.1 Temperature

In order to analyse the monthly maximum, minimum and mean temperature across the year, the daily temperature we have must be summarised by region and month. Maximum temperature is the maximum of all the maximum temperatures given in the region at any particular month. Minimum temperature is the minimum of all the minimum temperatures at any particular month. Similarly, the mean temperature is the mean of the mean temperatures given in the region at any particular month.

3.2.4.2 Rainfall

Firstly, the total rainfall for each station and each month is calculated. Secondly, the standard deviation (SD) of this monthly total rainfall is calculated. Lastly, the total rainfall for the month, the upper bound of the whisker (mean + 1 SD) and the lower bound of the whisked (mean - 1 SD) are calculated.

4 User Interface Design

4.1 Calendar Heatmap

Based on our prior research, we found some R packages that we could leverage on to visualise the weather data more appropriately. For the visualisation of annual temperature, we decided to narrow down to the mean value of the temperature for all stations and show it as one "square box" in the calendar, which represents 1 day. The calendar heatmap allows easy comparison of temperature/rainfall trends across different years.

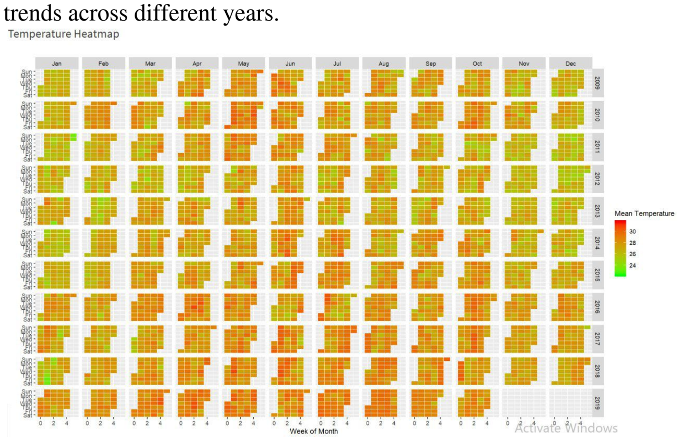

Figure 5

As seen from figure 5, annual temperature is now visualised in the form of daily mean temperature of Singapore across different stations, with the colour green representing lower temperature range and red representing higher temperature range. Such colour scheme was chosen to better differentiate the hotter days from cooler days.

4.2 Ridgeline Plot

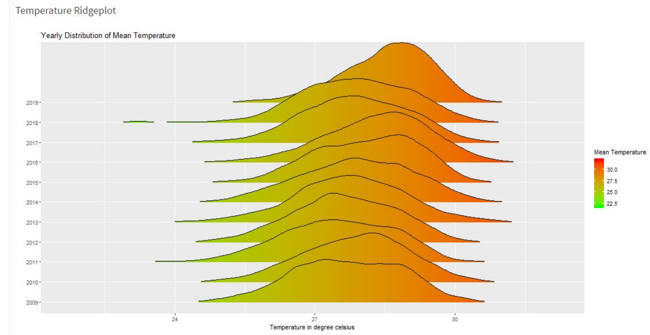

Figure 6

Figure 6 is an example of the ridgeline plot that was created with the use of package called ggridges.[4] Each line plot allows users to visualise the distribution of daily mean temperature/ daily rainfall from 2009 to 2019 (in descending order of the year in the figure). Users will be able to see the range of daily mean temperature/daily rainfall for each year and the peak of the line plot shows the temperature/rainfall where most days in the year are.

4.3 Geospatial Interpolation

Figure 7

Figure 7 is a year filter which applies to both Figure 8, 9 and Figure 10. This filter allows users to choose the year they wish to analyse.

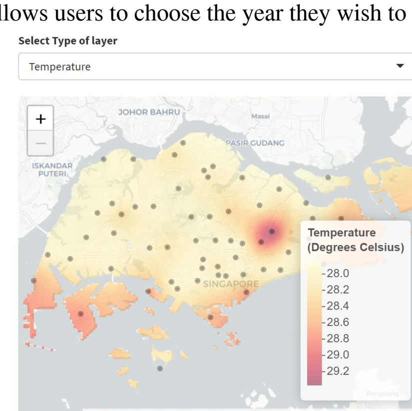

Figure 8

Geospatial interpolation is created by Raster package to display the raster image on leaflet to visualise the interpolated spread of temperature/rainfall. The circle markers within the map visualisation represents the station locations in Figure 8. There is a filter that allows users to choose the different layers of the interpolation of temperature/rainfall.

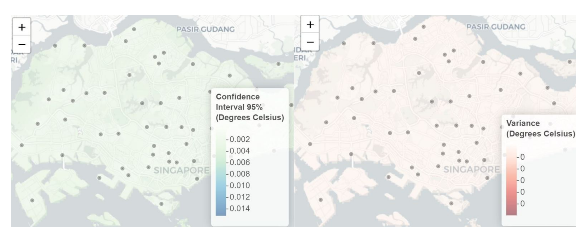

Figure 9

The filter in Figure 7 will be available for users to toggle to variance and confidence interval at 95% plots to address any uncertainty in temperature/rainfall data that has been interpolated with Gstat package as shown in Figure 9.

4.4 Monthly Distribution Across A Year

4.4.1 Temperature

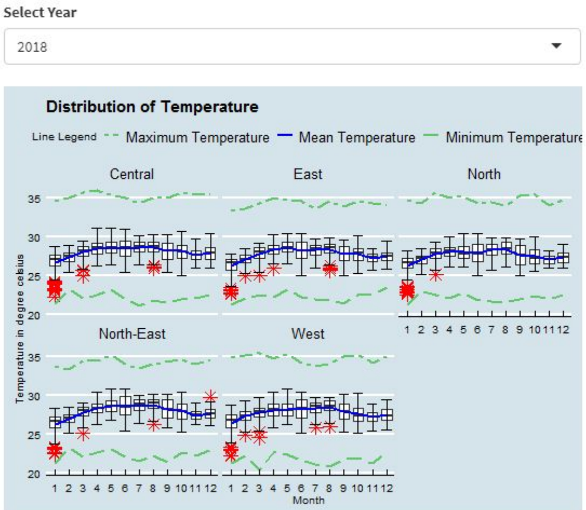

Figure 10

The filter for year in Figure 7 applies to this plot in Figure 10 as well. This plot is made up of boxplot and line graphs. This plot consists of boxplot for each month and outliers are marked in red. The boxplot shows the distribution of mean temperature across different regions for each month. The line graphs show the maximum, minimum and mean for the year by regions. These line graphs allow users to see the changes in maximum, minimum and mean of the temperature over the year.

4.4.2 Rainfall

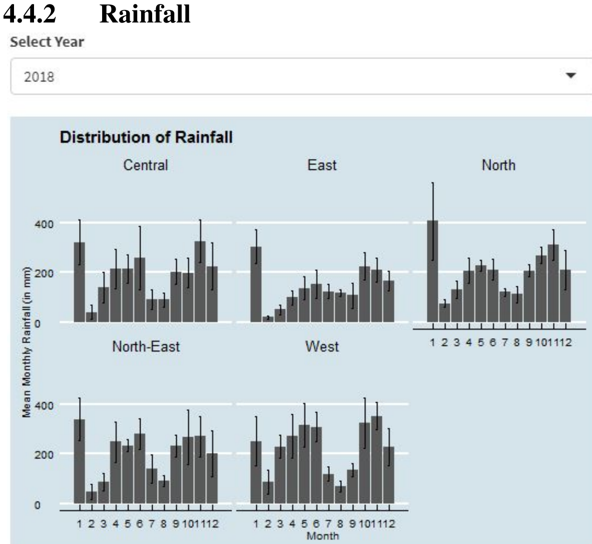

Figure 11

In figure 11, the rainfall distribution is presented using bar graph with standard deviation error bars. The bar graph represents the mean monthly rainfall across all stations in the respective regions.

The error bar shows the variance for the mean monthly rainfall between stations in each month. A short error bar means that the variance is small and hence, stations in that region experience narrow range of total rainfall for that month. Therefore, the mean monthly rainfall is more representative of all stations in that region. Conversely, if the error bar is long, it shows that the variance for that month in that region is high. Hence, the stations for that month have total rainfall that differ significantly from each other and thus, the bar for that month might not be representative of all stations in that region.

5 Web Application Architecture

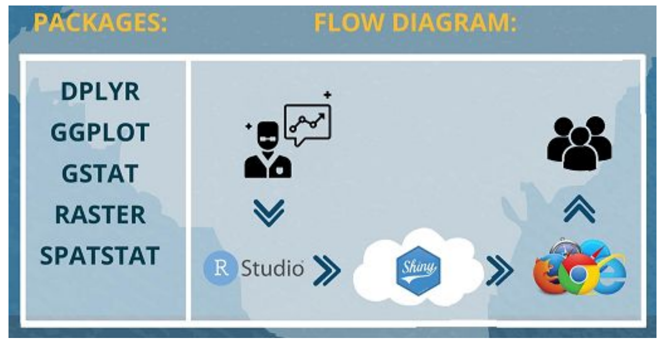

Figure 12

Figure 12 is a summary of our architecture. The code was tested and deployed locally through Rstudio's shiny app. For it to be publicly accessible, the application is deployed onto shiny cloud.

6 Results

6.1 Analysis of Singapore Temperature

6.1.1 Calendar Heatmap

As shown in Figure 4, the calendar heatmap tells us that the number of warmer months is gradually increasing. More specifically, the range of warmer months is becoming wider by year. For example, in year 2009, the warmer period was during May to August, as these months are more concentrated with orange/red than the rest of the year. However, the months with high concentration of orange/red gradually increase down the year.

Another interesting finding that we could visualise from the plot is in year 2018. In January 2018, there were a few days when the temperature decreased to as low as 24 degree celsius to the point that many of the SMU students were carrying a hoodie or jacket around school, and this is shown as a very green “box” if one were to take a closer look at January 2018.

6.1.2 Ridgeline Plot

With reference to Figure 6, we can see that the mean temperature is generally shifting towards the right from 2009 to 2019 (year in descending order from the top) with the exception of 2017 and 2018. The range of mean temperature has also decreased while the minimum mean temperature generally increased from 2009 to 2019.

The lowest mean temperature can be found in 2018. This probably corresponds to the five-day cool spell Singapore has experienced back in January 2017 [5].

6.1.3 Geospatial Interpolation

Based on Figure 8, it is evident that areas between central and east regions such as Paya Lebar experienced higher temperature, especially so from 2014 to 2019. Similar trend could be observed for areas in the southwest region in Singapore such as Tuas. Interestingly, these areas are where industrial buildings are commonly situated.

By studying the variance and confidence level at 95% in Figure 7, there is not much interpolated data that can be considered as extreme outliers. One possible reason for this could be that Singapore's land size is too small and does not have four seasons to observe wider temperature range.. Additionally, Singapore has a lot of weather stations which largely contribute to the high accuracy of the interpolations. Monthly Distribution Across A Year
With reference to Figure 10, the outliers are in general 1.5 times the interquartile range from the lower hinge of the boxplot. The distribution of the mean temperature generally does not differ much between within regions across all years. There are more outliers in January as compared to other months. This may suggest that there are more cooler days in January. However, as this is based on 2018, this finding corresponds to the month when we experienced cooler weather.
The line graphs generally change similarly across all regions and all years. The difference between regions is very small in Singapore, which is about 1 to 2 degree celsius. For instance, in Figure 10, the maximum temperature in the east and north-east is about 1 to 2 degree celsius lower than the other regions for the same month period.

6.2 Analysis of Singapore Rainfall

6.2.1 Calendar Heatmap

Figure 13

Figure 13 shows the annual total rainfall across stations in Singapore. However, there is no distinctive trend in the total rainfall in Singapore as compared to the heatmap for temperature. This could be due to the fact that Singapore is a country that experiences rain all year round, but the total rainfall does not vary to a large extent. There are some outliers, like the "boxes" that are shaded with darker blue, but are minimal.

Another finding that we could see from figure 4 is that the months when there is higher frequency of rain are usually November and December, and this can be applied to any of the years.

6.2.2 Ridgeline Plot

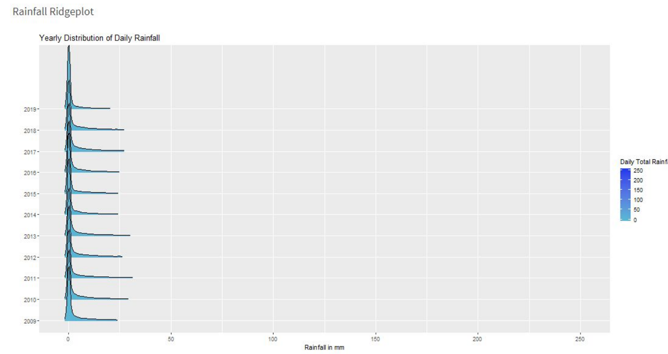

Figure 14

In Figure 14, the ridgeline plot of rainfall shows that the daily rainfall in Singapore is generally below 50mm. However, the x axis has shown that there are a few outliers with rainfall of more than 250mm resulting in the wide range. The amount daily rainfall is similar across all years.

6.2.3 Geospatial Interpolation

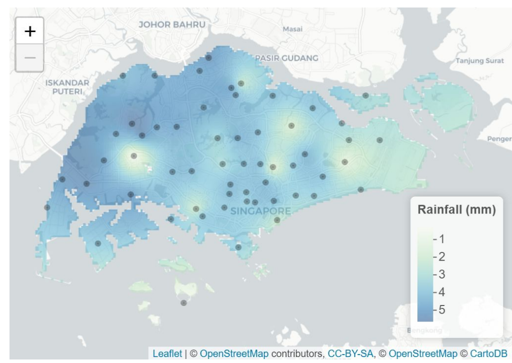

Figure 15

As shown in Figure 15, in the past few years dating back from 2014 to the present, it can be seen that the west areas experience higher rainfall. However, areas in the east seem to experience less rainfall as compared to the rest of the country,

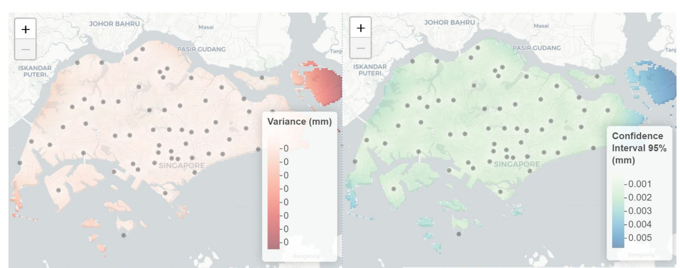

Figure 16

By studying the variance and confidence level at 95%, there is not as much interpolated data that can be considered as extreme outliers for the same reasons on analysis used for geospatial interpolation analysis for temperature as shown in Figure 16.

Hence, the only areas with increased rainfall level will be Pulau Tekong which is in the west of Singapore as the interpolation of data within that island was done with values from the weather stations in the mainland.

6.2.4 Monthly Distribution Across A Year

With reference to Figure 10, all the regions have very similar trend in the rainfall across the year. Throughout the year, different regions have different month(s) where each region is lower or higher than the others. However, in general, the east region seems to have consistently lower rainfall as compared to the other regions.

The error bars are generally longer when there was higher rainfall and shorter when the amount of rainfall was lower. A longer error bar suggests that there is higher variance in the total amount of rain in that month across stations. Hence, this may suggests that some stations in that region experience high amount of rain while some stations in that region experience low amount of rain resulting in a higher variance. Therefore, the mean total amount of rain for the month is not representative of all stations in that region.

7 Limitations

1. The data that we used is historical data, so our application does not show any real-time analysis of the Singapore climate.
2. The scope of our project had to be narrowed to temperature and rainfall despite the fact that we had wind data. This was because the wind data only gives us the speed, and not its direction so the insights gained would be minimal.
3. The package that could have been easily used to normalise the week variable for calendar heatmap (ddply)) wasn't available for the latest version of R, so we had to use the week.of.month() function in a for-loop which increased the loading time when publishing the dashboard.
4. 2019 data excludes data in November and December hence, analysis for the year 2019 will not be accurate as these data are missing and can be significant since these two months are considered the rainy season in Singapore.

8 Conclusion

Through our analyses of rainfall and temperature, we conclude that the mean temperature has definitely increased over the 10 years showing the impact of global warming on Singapore. Although, the change is gradual, we recommend that everyone should start putting in more individual effort to tackle this problem before it is too late.

Secondly, it is useful to note that the hotter areas in Singapore are coincidentally industrial area and its surroundings. As such, this can be a consideration for people who are looking for houses or businesses which temperature can be a concern. For instance, businesses that sell cold desserts might be interested in setting up their businesses in these areas.

Lastly, the east region tends to have less rain and has generally slightly lower temperature than the rest of the regions. This is in contrast with a common stereotype that it will be cooler if there is more rain. However, this was not inspected in this research and we did not establish whether there is any correlation between rainfall and temperature and hence, this can be looked into for future study.

9 Acknowledgement

The Climate Vizards would like to show our utmost gratitude and respect to Prof Kam Tin Seong for his endless support and guidance that helped us head towards the right direction whenever we faced difficulties or challenges throughout the semester. We will never forget those times when he would come over to help us with R even when he was on the way home.

References

[1] National Day Rally 2019 on Climate Change: https://www.channelnewsasia.com/news/singapore/ndr-2019-climate-change-impact-singapore-greatest-threat-sea-11819382
[2] Historical Daily Records: http://www.weather.gov.sg/climate-historical-daily/
[3] Geospatial Interpolation: https://mgimond.github.io/Spatial/interpolation-in-r.html
[4] Ggridge: https://cran.r-project.org/web/packages/ggridges/vignettes/introduction.html
[5] Five-day Cool Spell: https://www.straitstimes.com/singapore/environment/five-day-cool-spell-was-singapores-longest-in-a-decade