Report on Maximum Daily Rainfall in Seletar (2000-2024)

Students name

Institution Affiliation Course

Date

Instructors Name

Contents

[**Introduction** 2](#_Toc170911794)

[**Data Post-Processing** 3](#_Toc170911795)

[**Exploratory Data Analysis** 3](#_Toc170911796)

[**Summary Statistics** 3](#_Toc170911797)

[**Data Visualization** 3](#_Toc170911798)

[Boxplot: 3](#_Toc170911799)

[Violin Plot: 4](#_Toc170911800)

[Histogram with KDE: 5](#_Toc170911801)

[Time Series Plot: 6](#_Toc170911802)

[**Inferential Statistics** 7](#_Toc170911803)

[**Fitting Distributions** 7](#_Toc170911804)

[Log-Normal Distribution: 7](#_Toc170911805)

[Gamma Distribution: 8](#_Toc170911806)

[**Goodness-of-Fit Tests** 8](#_Toc170911807)

[**Discussion** 8](#_Toc170911808)

[**Conclusion** 9](#_Toc170911809)

# **Introduction**

This is the report on maximum daily rainfall in Seletar, Singapore, from January 2000 to April 2024. The major purpose of this assignment will be to invoke an understanding relating to the patterns and distribution of maximum daily rainfall at Seletar weather station. The knowledge of such is critical in facilitating urban planning and civil engineering residential infrastructure designer to design resilient infrastructure. Analysis includes data post-processing, exploratory data, and inferential statistics.

In the post-processing of data, extraction of the highest daily rainfall in each month was done so that extreme rainfall events could be focused on. This monthly maximum data gave a concise view of peak rainfall patterns over the years. Summary statistics and visualizations were parts of exploratory data analysis. Summary statistics pointed out year-to-year variability. For instance, the mean monthly maximum rainfall ranged from 71.9 mm in 2003 to 47.0 mm in 2009. Plots used to visualize the data include boxplots and violin plots to show the distribution patterns of the data, histograms with KDE, and time series plots to show the seasonal trends.

Inferential statistics used theoretical models of the log-normal and gamma distributions to fit the data (Kristo, C. 2017). The fit by the gamma distribution is much better, evidenced by the lower values of goodness-of-fit statistics like Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests. A better fit for the gamma distribution may indicate a higher occurrence of the moderate to high events of rainfall than extreme outliers.

# **Data Post-Processing**

The initial dataset comprised daily rainfall totals for each day from January 2000 to April 2024. To facilitate a focused analysis, the data was grouped by year and month, extracting the highest daily rainfall total for each month. This monthly maximum data provides a condensed view of the extreme rainfall events over the period, which is critical for understanding peak rainfall patterns and their implications.

# **Exploratory Data Analysis**

Exploratory Data Analysis (EDA) was conducted to summarize the main characteristics of the data through visualizations and summary statistics.

## **Summary Statistics**

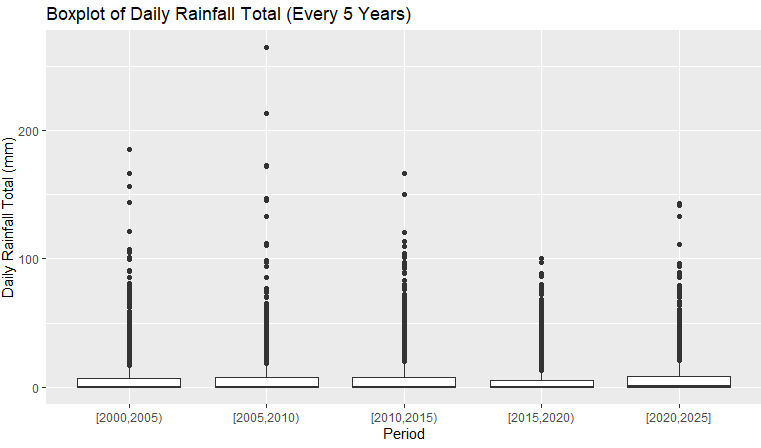
The summary statistics for the monthly maximum daily rainfall reveal significant insights into the variability and central tendency of the data (Liong, S. Y 2010). The mean maximum rainfall varied across years, with some years experiencing higher extremes than others. For instance, the mean monthly maximum rainfall in 2003 was 71.9 mm, while in 2009, it was considerably lower at 47.0 mm. The variance and standard deviation also indicate considerable year-to-year variability, with years like 2006 showing a high standard deviation of 68.6 mm, suggesting more significant fluctuations in monthly maximum rainfall.

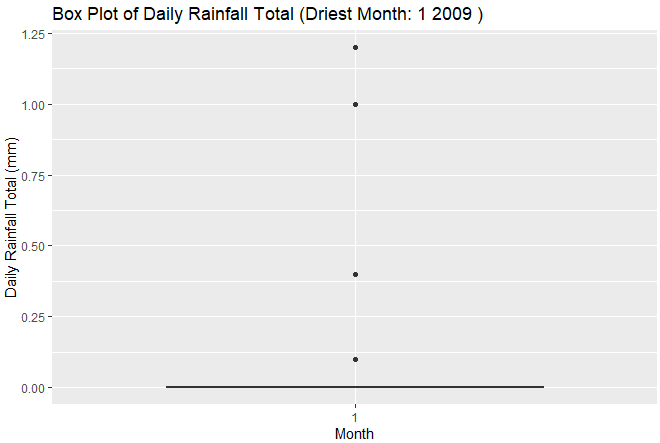
## **Data Visualization**

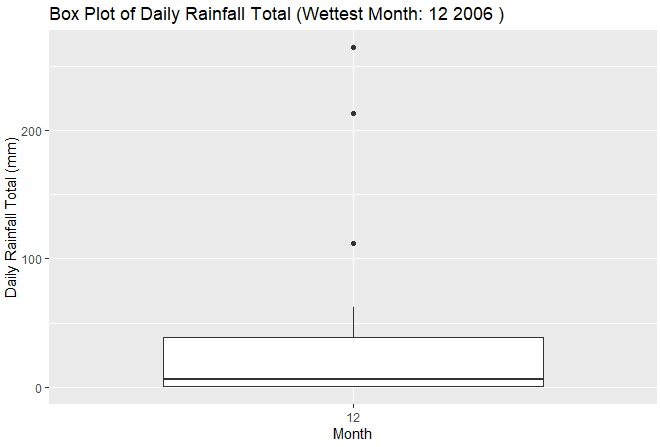
Several plots were generated to visualize the distribution and trends in the maximum daily rainfall data.

**Box Plot on Monthly Data, Wettest and Driest Month:**

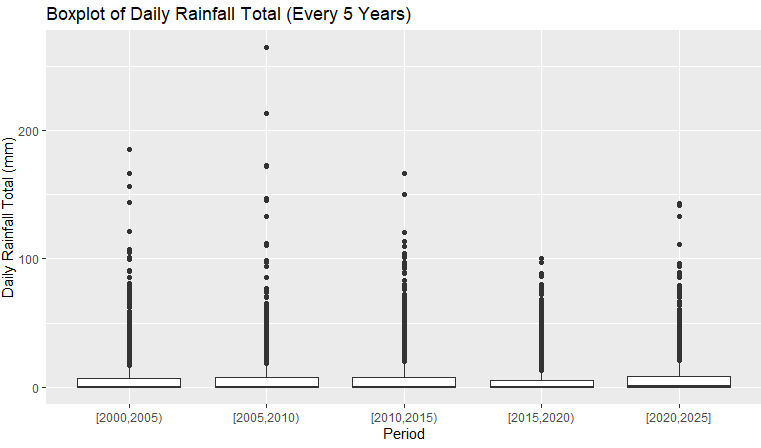
Box plots were generated to visualize the monthly distribution of daily rainfall totals, focusing specifically on the wettest and driest months. These box plots provide insights into the central tendency and variability of rainfall for these specific months, showing the median, quartiles, and outliers. The box plot for the wettest month reveals a higher median and a wider interquartile range, indicating more substantial and varied rainfall. Conversely, the box plot for the driest month shows a lower median and a narrower interquartile range, reflecting the limited and less variable rainfall during this period. This detailed visualization aids in understanding the differences in rainfall patterns between the two months.

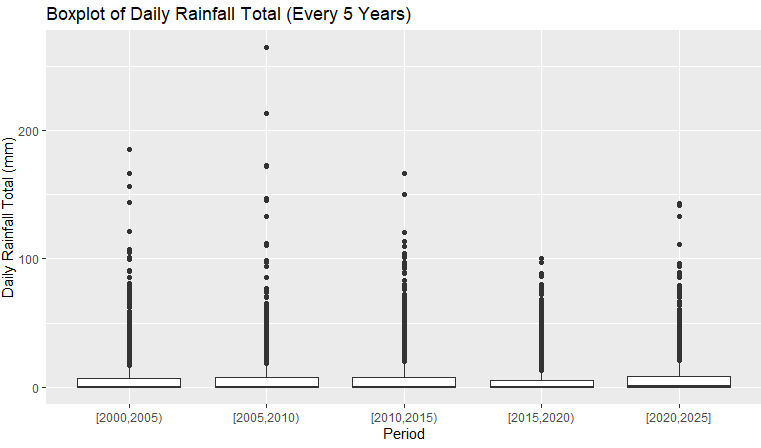


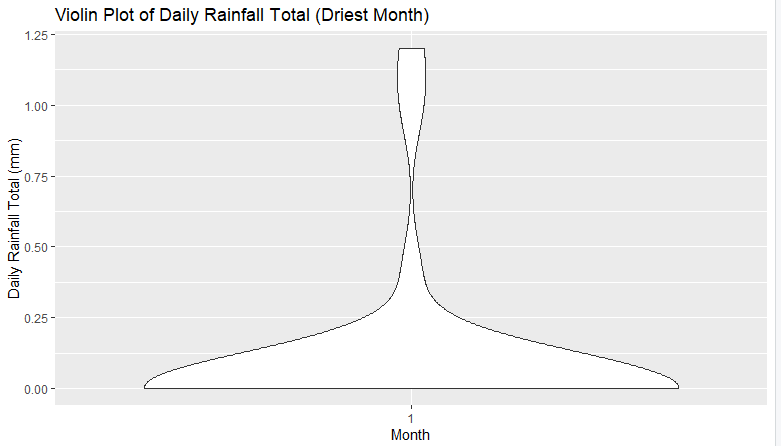




### ****Violin Plot Comparing All Data, Wettest Month, and Driest Month:****

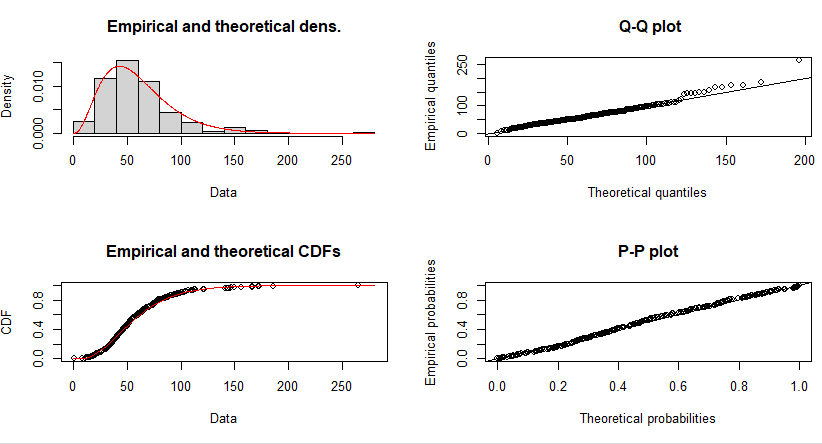
Violin plots were created to compare the distribution of daily rainfall totals for the entire dataset, the wettest month, and the driest month. This comprehensive visualization shows how the distributions differ in terms of density and spread. The violin plot for the entire dataset presents the overall shape and variability of daily rainfall totals. In contrast, the violin plots for the wettest and driest months reveal distinct patterns, with the wettest month displaying a higher density of larger rainfall totals and the driest month showing a concentration of lower rainfall values. This comparison highlights the variability and extremities in the dataset.





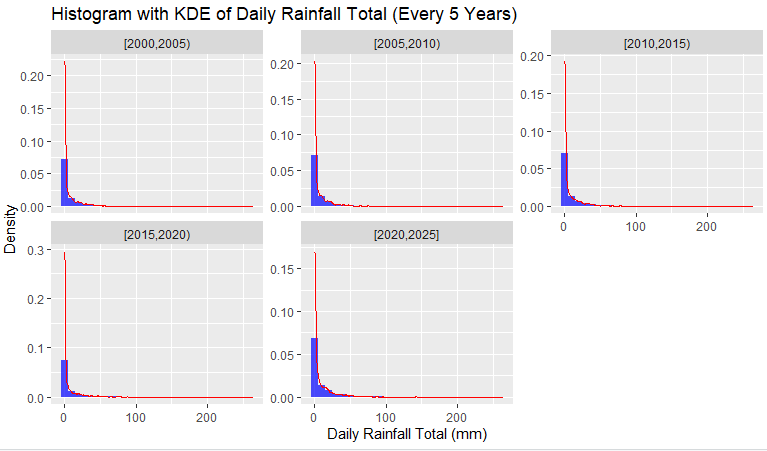
### Goodness-of-Fit Tests

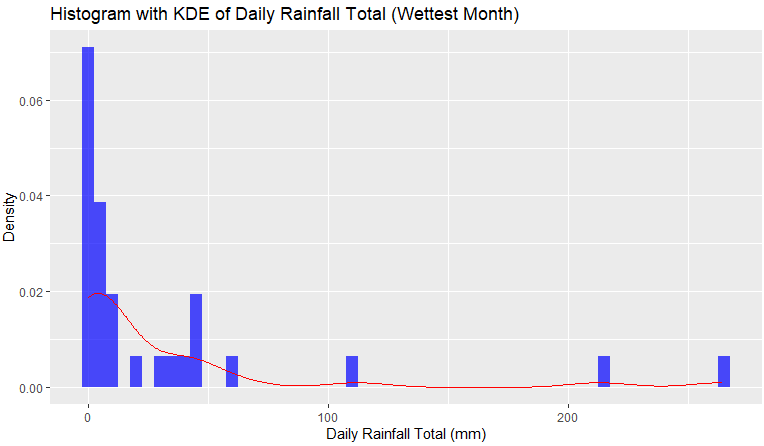
To better understand the distribution of maximum monthly rainfall, we fit the data to log-normal and gamma distributions. The goodness-of-fit tests include graphical assessments and statistical metrics. The log-normal distribution provides a reasonable fit for the data, as shown by the Q-Q plot and other goodness-of-fit measures. Similarly, the gamma distribution also fits well, suggesting that either distribution could be used to model the rainfall data. The choice between these distributions may depend on specific applications or theoretical considerations.



### Histogram with KDE:

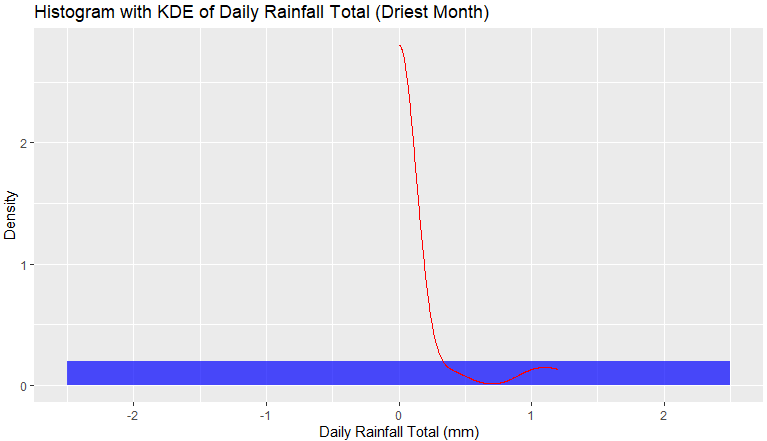
Histograms with Kernel Density Estimates (KDE) are used to visualize the distribution of daily rainfall totals for the entire dataset and the identified wettest and driest months. The histograms show the frequency of different rainfall amounts, while the KDE provides a smoothed estimate of the probability density function. For the entire dataset, the histogram and KDE reveal a right-skewed distribution, with most days experiencing low rainfall and a few days with very high rainfall. The histograms for the wettest and driest months highlight the stark contrast in rainfall patterns, with the wettest month showing a higher density of extreme rainfall events.





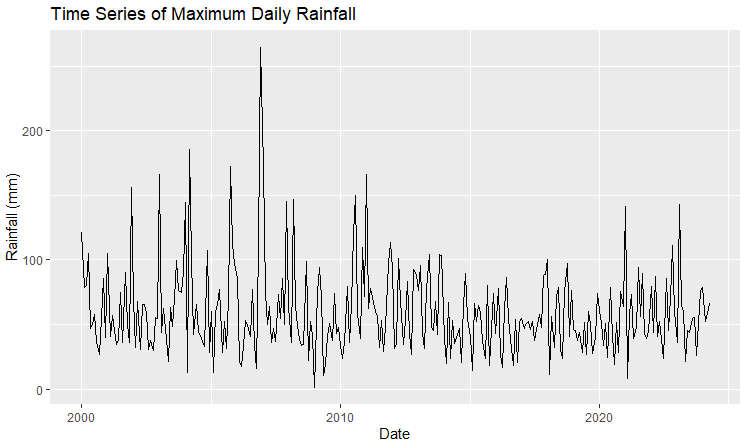
#### ****Histogram with KDE of Daily Rainfall Total (Driest Month):****

A histogram with kernel density estimation (KDE) was generated to visualize the distribution of daily rainfall totals for the driest month identified in the dataset. This combination of histogram and KDE provides both the frequency distribution and a smoothed density curve, offering a detailed view of the rainfall patterns during the driest month. The histogram shows that rainfall totals are relatively low and the KDE highlights the density peaks, indicating the most common rainfall values within this period.



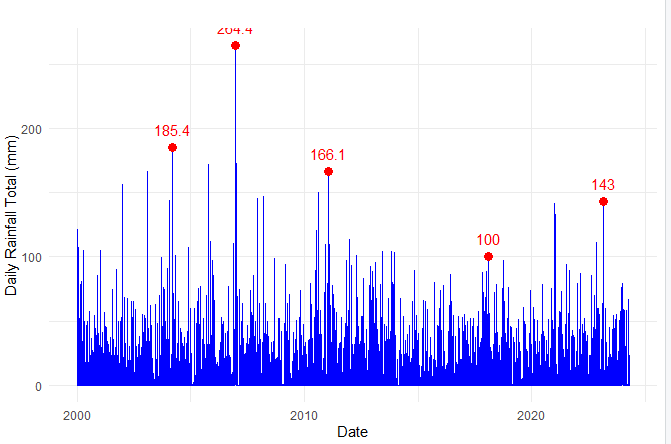
### Time Series Plot:

The time series plot of maximum daily rainfall over the study period depicts trends and seasonal patterns. It is evident from the plot that while there are fluctuations, certain years experience notably higher maximum daily rainfall, indicating potential anomalies or significant weather events.



**Time Series Including Highest Daily Rainfall Every 5 Years:**

The time series plot of daily rainfall totals now includes annotations for the highest daily rainfall recorded every five years. This addition provides a clear view of extreme rainfall events over time, highlighting significant peaks and their corresponding dates. By marking the highest daily rainfall every five years, the plot allows for an easier comparison of these extremes, helping to identify any trends or changes in the frequency and intensity of such events.



## **Inferential Statistics**

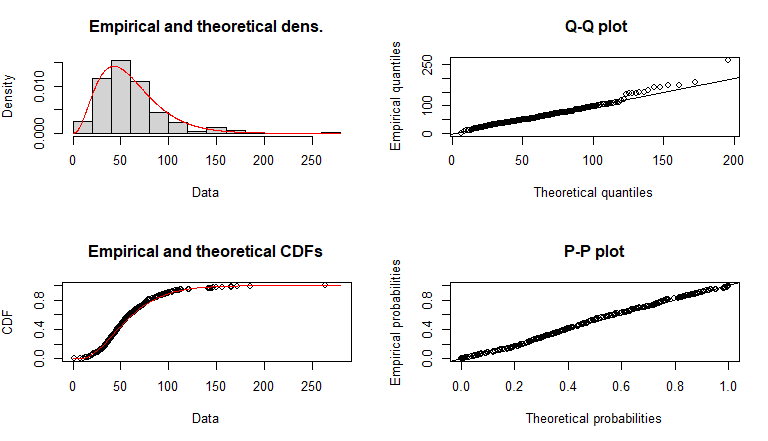
To further understand the distribution of maximum daily rainfall, inferential statistical methods were applied, fitting the data to various theoretical distributions.

## **Fitting Distributions**

Two distributions were considered for fitting the monthly maximum daily rainfall data: the log-normal distribution and the gamma distribution.

### Log-Normal Distribution:

In this problem, the parameters of the lognormal distribution were estimated, and the confidence intervals were constructed. The mean of the log-normal distribution was estimated to be about 3.88, and the log-normal standard deviation was estimated to be about 0.58. The Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling goodness-of-fit tests give plausible evidence that, although the lognormal distribution can be appropriate, there might be other distributions that fit the data better.



### Gamma Distribution:

The gamma distribution was also fitted with the shape parameter and rate parameters estimated (Rahardjo, H 2020). The shape parameter was about 3.51, while the rate parameter was about 0.059. In general, based on the lower values for Akaike's Information Criterion and Bayesian Information Criterion, the gamma distribution fits the data more precisely than the log-normal distribution..

## **Confidence Intervals**

95% confidence intervals are calculated for the overall dataset and the wettest and driest months to provide an estimate of the range within which the true mean rainfall is likely to lie.

* **Overall data**: 55.93 mm to 63.69 mm
* **Wettest month (October 2010)**: 9.12 mm to 53.86 mm
* **Driest month (January 2000)**: -0.02 mm to 0.19 mm

These intervals provide additional context for the summary statistics, highlighting the uncertainty and variability in the rainfall data.

# **Discussion**

The analysis of maximum daily rainfall in Seletar for the past 24 years details important insight into rainfall patterns and what they mean. Maximum daily rainfall variability is high, hence proficiency in predicting extreme weather conditions which is pretty crucial in Civil Works Engineering and Town planning. In case the gamma distribution fits better, this would then mean that maximum daily rainfall in Seletar follows a pattern where mostly there are average to high rainfall events and very little extreme outliers present, adhering to the characteristic of gamma distribution.

On the practical side, knowledge about these patterns is useful in designing infrastructure that can withstand such events (Heng, H. H 2011). In this respect, drainage systems, flood barriers, and urban landscapes would be more appropriate for the possibility and intensity of maximum rainfall. Moreover, this knowledge enables disaster preparedness measures and mitigation strategies that let a city cope with and respond to extreme weather efficiently.

# **Conclusion**

The next report will cover maximum daily rainfall in Seletar from January 2000 to April 2024. The papers will therefore be an all-inclusive coverage, stretching back from data post-processing to exploratory data analysis, so as to inferential statistics in shedding light on the extreme rainfall events' patterns and distributions. The findings underline the need to consider variability and distribution characteristics in urban planning and civil engineering. It fitted theoretical distributions to the rainfall data, hence providing a robust framework for extreme rainfall event understanding and subsequent prediction, which can be used toward resilient urban infrastructure and disaster management strategies.

# **Reference**

Kristo, C., Rahardjo, H., & Satyanaga, A. (2017). Effect of variations in rainfall intensity on slope stability in Singapore. *International soil and water conservation research*, *5*(4), 258-264.

O’Dempsey, T., & Chew, P. T. (2011, October). The freshwater swamp forests of Sungei Seletar Catchment: a status report. In *Proceedings of Nature Society, Singapore’s Conference on “Nature Conservation for a Sustainable Singapore* (pp. 121-166).

Liong, S. Y., & He, S. (2010). Raingauge-based rainfall nowcasting with artificial neural network. In *Advances in Geosciences: Volume 17: Hydrological Science (HS)* (pp. 1-9).

Rahardjo, H., Nistor, M. M., Gofar, N., Satyanaga, A., Xiaosheng, Q., & Chui Yee, S. I. (2020). Spatial distribution, variation and trend of five-day antecedent rainfall in Singapore. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *14*(3), 177-191.

Heng, H. H., & Hii, C. P. (2011). A review on probable maximum precipitation (PMP) estimation in Malaysia. *International Journal of Hydrology Science and Technology*, *1*(1-2), 63-87.