WQD7001 Principle of Data Science

Session 2023/2024: Semester 1

Group 5

Title: Predictive Analysis for Precipitation in Kuala Lumpur

Lecturer: Dr Maizatul Akmar binti Ismail

Name	Matric Number
CHEN XIN WEI	17202767
SUM SI YENG	22068843
TAN JIA HUI	22113856
TAN WEI VEN	22107624
YEAP YU XIAN	17203286

Table of Contents

1.0 Project Background

Climate change poses significant challenges globally, impacting various facets of our environment and necessitating innovative solutions. One critical aspect that demands attention is the accurate prediction of rainfall, especially in regions prone to weather extremes. In this context, Kuala Lumpur, which is a densely populated urban center, emerges as a pivotal location for the focus of this project. Given its tropical climate and susceptibility to heavy rainfall, the city serves as a crucial case study for enhancing our understanding and predictive capabilities.

As we witness the intensification of climate change, the need for precise and timely rainfall predictions becomes even more urgent to effectively address the challenges that may arise. The urban landscape of Kuala Lumpur faces challenges related to heavy rainfall, including flash floods and infrastructure strain. Accurate rainfall prediction plays a crucial role in disaster preparedness and response, enabling authorities to implement timely measures to safeguard public safety and mitigate the impact on critical infrastructure.

In light of these considerations, this project aims to contribute to the advancement of rainfall prediction methodologies in Kuala Lumpur. This study also aims to empower stakeholders across various sectors, from environmental conservation to urban planning, with insights that foster resilience in the face of changing weather patterns.

2.0 Problem Statement

Traditional methods of rainfall prediction often fall short in providing accurate and timely information, especially in the context of Kuala Lumpur, Malaysia. The region's unique climate dynamics, influenced by factors such as monsoons and localized weather patterns, contribute to the complexity of predicting precipitation events. Inaccurate forecasts can have far-reaching consequences, affecting both businesses and the general public, necessitating a more sophisticated approach to predictive analysis.

The increased frequency and intensity of extreme weather events, driven by climate change, pose risks to various sectors and communities. Sudden and heavy rainfall can lead to floods, resulting in property damage, disruptions to transportation networks, and financial losses. These adverse impacts extend beyond businesses to impact the daily lives of residents, emphasizing the need for a predictive analysis framework for rainfall in Kuala Lumpur to implement adaptive strategies and enhance resilience.

In response to these challenges, the project aims to enhance rainfall prediction methodologies through advanced predictive analytics tailored to the specific characteristics of Kuala Lumpur's climate. By leveraging data-driven approaches, the goal is to provide timely and accurate information not only to businesses but also to the public.

Through this predictive analysis on precipitation, the project seeks to contribute to minimizing the adverse and disruptive effects caused by changes in weather patterns and climate in Kuala Lumpur. The project aims to foster a more resilient and prepared community in the face of unpredictable rainfall patterns.

3.0 Project Objectives

Below are the objectives identified for the project:

1. To Identify the Optimal Machine Learning Model for Rainfall Prediction in Kuala Lumpur.
2. To evaluate and compare the performance of selected machine learning algorithms in predicting rainfall in Kuala Lumpur.
3. To evaluate various metrics to determine the most suitable evaluation criteria for assessing the performance of the machine learning model in the context of rainfall prediction in Kuala Lumpur.

4.0 Project Scope / Domain

The project resides within the domain of environmental studies, with a primary focus on developing rainfall forecasting methodologies using machine learning techniques. Specifically, the scope of this study is confined and tailored explicitly to the characteristics of Kuala Lumpur's climate.

The primary goal is to develop machine learning models specifically designed for the unique characteristics of Kuala Lumpur's climate. This model seeks to outperform existing forecasting methods, providing a more accurate and dependable tool for predicting rainfall patterns in Kuala Lumpur. By concentrating on Kuala Lumpur, this project aims to tackle the challenges in the local weather patterns. Advanced analytics and machine learning techniques will be employed to gain comprehensive understanding of Kuala Lumpur's climate and rainfall patterns.

5.0 Literature Study

Table 1.1 below includes six research articles published in recent years, which are deemed as valuable resources for our model development and the formulation of project objectives. A thorough review has been conducted, summarizing the work and key findings of each article.

No.	Authors	Title	Summary of work and key findings
1.	Sulaiman, N. A. F., Shaharudin, S. M., Ismail, S., Zainuddin, N. H., Tan, M. L., & Jalil, Y. A. (2022)	Predictive Modelling of Statistical Downscaling Based on Hybrid Machine Learning Model for Daily Rainfall in East-Coast Peninsular Malaysia	• This study focuses on Kelantan and Terengganu states in Peninsular Malaysia, which are situated on the east coast. The study incorporates 10 strategically selected rainfall observation stations across the chosen region. • Through Principal Component Analysis (PCA), the study effectively reduced the high-dimensional dataset to six principal components, capturing over 70% of the total variance. The careful selection of predictors based on PCA loadings ensured that the most influential factors were retained for further analysis. This approach streamlined the dataset and also facilitated the identification of factors significantly influencing rainfall patterns in east-coast Peninsular Malaysia. • This study evaluated various machine learning models for statistical downscaling, with Support Vector Classification (SVC) and Support Vector Regression (SVR) demonstrating solid performance. However, the hybrid model of Support Vector Classification and Relevant Vector Machine (SVC-RVM) emerged as the most robust approach. This model not only outperformed

			others in terms of accuracy, misclassification error, and support vectors but also excelled in forecasting daily rainfall. The SVC-RVM hybrid model showcased a superior ability to predict rainfall patterns accurately for the next two years, making it a recommended choice for statistical downscaling in the context of climate change impact assessment.
2.	Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022)	Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting	• This study compared XGBoost, AutoML ensemble, and LSTM-based models for hourly rainfall prediction in major UK cities. • The AutoML ensemble, combining Gradient Boosting, Linear Support Vector, and Extra-trees Regressors, outperformed XGBoost, with Cardiff showing the best overall performance. LSTM and Stacked-LSTM models consistently demonstrated superior predictive abilities compared to traditional machine learning approaches. • This study suggested that model complexity played a crucial role in performance, as evidenced by the poor performance of a 10-layer Stacked-LSTM model compared to shallower networks. This study also emphasized the need for careful architecture selection and hyperparameter tuning.
3.	Liyew, C. M., & Melese, H. A. (2021)	Machine learning techniques to predict daily rainfall amount	• This study focuses on predicting daily rainfall intensity in Bahir Dar City, Ethiopia, utilizing machine learning algorithms. Environmental features collected from a

			meteorological station were analyzed for their relevance in rainfall prediction. • This study stands out for its meticulous analysis of environmental features' relevance through Pearson correlation. This approach ensures that only significant variables are utilized for predicting daily rainfall, enhancing the accuracy of the models. Relevant features, determined through Pearson correlation, were used as inputs for three machine learning models: Multiple Linear Regression (MLR), Random Forest (RF), and XGBoost. • This study suggested that XGBoost was found to be the most suitable algorithm for daily rainfall prediction, with the highest prediction accuracy.
4.	Mohammed, M., Kolapalli, R., Golla, N., & Maturi, S. S. (2020)	Prediction Of Rainfall Using Machine Learning Techniques	• This study focuses on offering non-experts easy access to the techniques, approaches utilized in the sector of precipitation prediction, especially in rainfall prediction as well as provide a comparative study among the various machine learning techniques which can be used to predict rainfall. • This study uses and compare three different regression analysis technique (Multiple Linear Regression (MLR), Support Vector Regression (SVR) and Lasso Regression) to predict rainfall from year 1901 to 2015 for each state in India. • This study concludes that Support Vector Regression (SVR) is a more valuable and adaptable strategy to be used. As SVR able

			to provide the lease errors in rainfall prediction compared to the other two methods.
5.	Adaryani, F. R., Jamshid Mousavi, S., & Jafari, F. (2022)	Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN	• This study focuses on short-term (5 minutes and 15 minutes ahead) rainfall prediction of Niyavaran station of Tehran, capital of Iran from 1974 to 2014 using three different machine and deep learning-based (PSO Support Vector Regression (PSO-SVR), Long-Short Term Memory (LSTM) and Convolutional Neural Network (CNN)) model. • This study indicates that PSO-SVR and LSTM approaches performed almost the same and better than CNN. Classification of the events will be able to improve the forecast models accuracy. • This study concludes that each models have its own strengths and weaknesses. In terms of short-term rainfall prediction, PSO-SVR and LSTM will be a better choice. However, which model is more suitable to be used still need to based on what are the outcome researchers want.
6.	Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2020)	Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia	• This study uses several machine learning models (Bayesian Linear Regression (BLR), Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR) and Neural Network Regression (NNR)) and methods (Forecasting Rainfall Using Autocorrelation Function (ACF) and Forecasting Rainfall Using Projected Error)

			to predict the rainfall data in 10 stations covering the Tasik Kenyir, Terengganu. • This study uses the Thiessen Polygon method to calculate the proximity area around the point with other points. It also can help to calculate station rainfall weight and average rainfall based on each station. • This study concludes that the results presented that for ACF gets better with cross-validation with BDTR and tuning its parameter. Also, the more input included to the model, the more accurate the model can perform.

Table 1: Summary for reviewed journal articles

6.0 Description of Methodology

6.1 Obtain

In the initial phase of our methodology, the meteorological data essential for rainfall prediction in Kuala Lumpur was acquired from the online weather data provider, Visual Crossing. This platform offers comprehensive and high-quality weather datasets, including information on precipitation, temperature, maximum and minimum temperature, temperature feeling, maximum and minimum temperature feeling, dew, humidity, precipitation coverage, cloud coverage, visibility, solar radiation, solar energy, un index and precipitation description. The decision to utilize Visual Crossing was guided by the platform's reliability, extensive coverage, and the availability of historical weather data specific to the Kuala Lumpur region. The dataset obtained from Visual Crossing encompasses a temporal scope that spans 1000 days, from 1st February 2021 to 28th October 2023 and consists of 1000 observations and 33 variables.

6.2 Scrub

From the original dataset, our initial step is removing redundant or irrelevant variables, such as name and datetime. After removing those variables, there are some categorical variables existing in our dataset. Before analyzing those variables, we transform them from characters into factors. After that, we proceed to examine the relationship between our dependent variable and independent variables to help us to identify those variables that are having insignificant relationship with the dependent variable and to remove them. In the last part of data scrubbing, we used to define missing value, outlier and duplicate value from datasets. In this part, we retained all existing data without removing any missing values or duplicates, as none were identified during our thorough examination. We do not remove outliers from our dataset. The rationale behind this decision is rooted in the belief that outliers, rather than being anomalies to be discarded from analysis, can hold valuable information in our study. In the context of rainfall prediction, extreme weather events or unusual patterns might be precisely what we need to capture and investigate, in order to build more robust and reliable models.

6.3 Explore

6.3.1 Basic Exploration

Dataset after scrubbing consists of 1000 observations and 16 variables. First round of basic data exploration was performed before data scrubbing to define the data that need to be cleanse. The exploration process involves function including names(), class(), dim(), str(), summary(), head() and tail() to see the overall data frame pattern. Dataset is found consisting of categorical and numerical variables. Therefore, we transformed the categorical variables into correct structure using as.factor() function.

6.3.2 Univariate Analysis

After selecting the variables that are significant to dependent variable, we proceed to perform univariate analysis for each of the variable to see their pattern and distribution. We used boxplot and histogram for numerical variable to identify trend and pattern. For categorical variable, we created bar chart using ggplot() to visualize the pattern.

As numerical variables, we observed that several variables such as tempmax, tempmin, temp, feelslikemax, feelslike, dew, humidity, precip, precipcover, visibility, solarradiation and solarevergy have outliners shown in the boxplot (refer to appendix) while cloudcover did not exhibit outliners. Based on histogram, tempmax, temp, feelslikemax, dew, humidity, solarradiation and solarenergy skewed to left. Conversely, tempmin, feelslikemin, feelslike, precip, precipcover, cloudcover and visibility shown rightward skeweness.

As category variables, the highest frequency of description is partly cloudy throughout the day (346) followed by partly cloudy throughout the day with rain (183). The lowest frequency is partly cloudy throughout the day with a chance of rain throughout the day (7). Next, the highest frequency of uvindex is 9 (315) followed by 8 (184) while the lowest frequency is 1 (3) (refer to appendix).

6.3.3 Bivariate Analysis

Correlation matrix was utilized to examine the relationship between dependent variable (precipitation) and independent variables. We also use scatter plots to represent the relationship between dependent variable and numerical independent variables. ANOVA test was used to determine the relationship between categorical variables and dependent variable. The correlation matrix below shows the correlation between every variable.

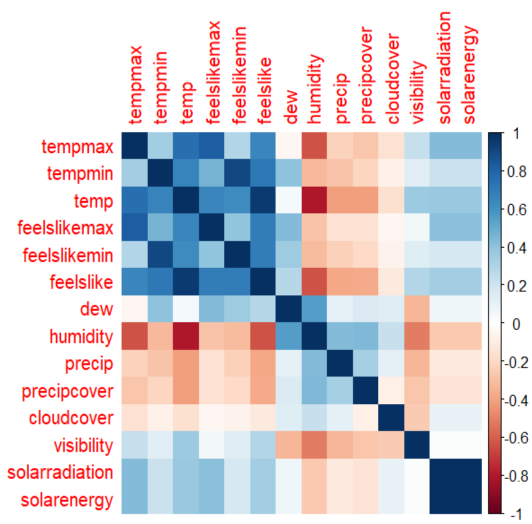


Figure 1: Correlation matrix

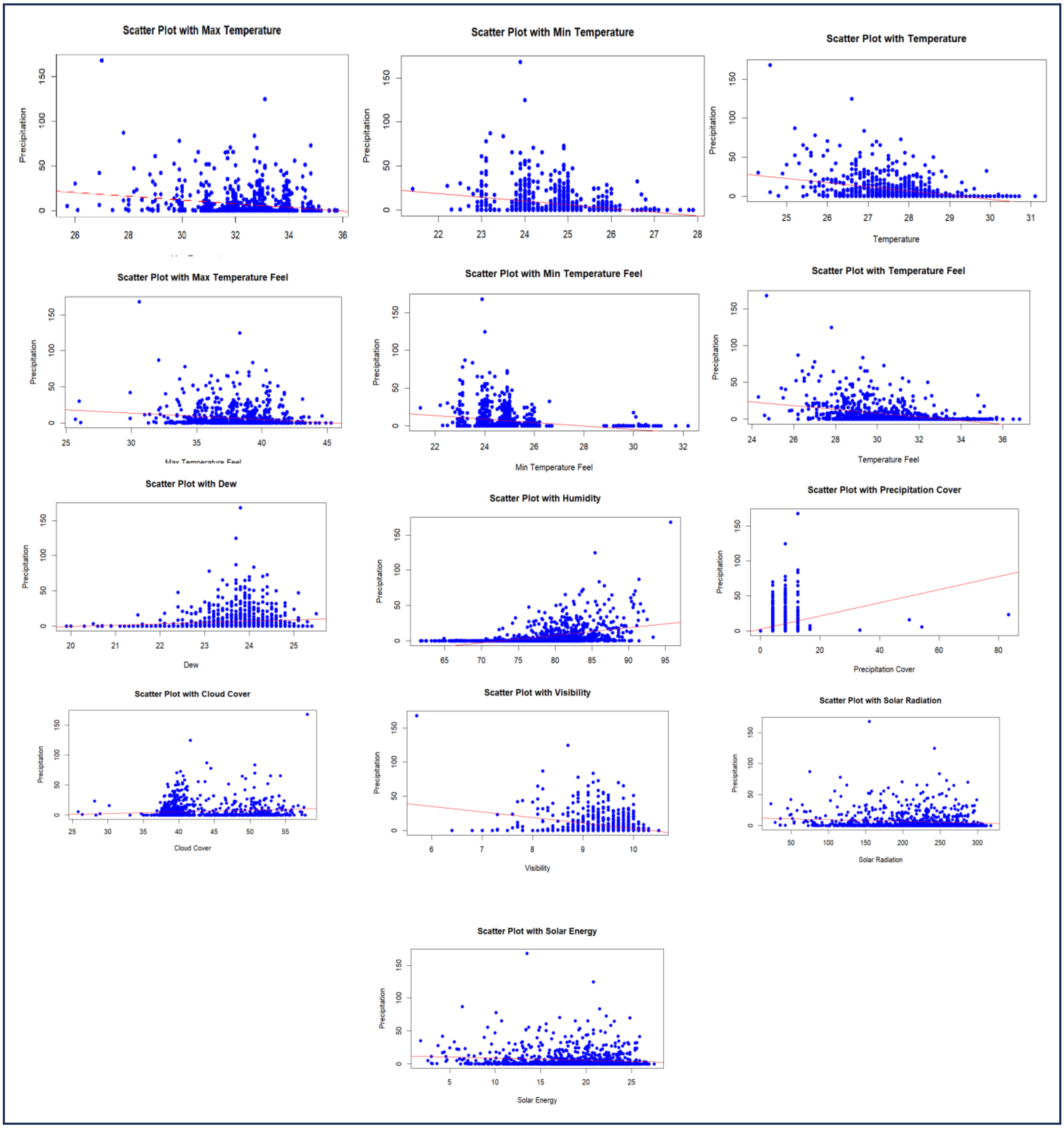


Figure 2: Scatter Plots for Independent Variables against Precipitation

According to the scatter plots, maximum temperature, minimum temperature, temperature, maximum temperature feel, minimum temperature feel, temperature feel, visibility, solar radiation, and solar energy show negative correlation with the precipitation. Dew, humidity, precipitation cover, and cloud cover are having positive correlation with precipitation.

```
                     Df Sum Sq Mean Sq F value Pr(>F)
description           8  39139    4892  30.606 <2e-16 ***
uvindex               9    825      92   0.573  0.820
description:uvindex  51   7079     139   0.868  0.732
Residuals           931 148822     160
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3: ANOVA test for categorical variables against precipitation

H_0 : All group means are equal.

H_1 : At least one group mean is different

From the ANOVA test result, we conclude that there is a significant difference in mean weight loss between description and precipitation since the p-value is less than 0.05. The main effects of uvindex and the interaction effect between description and uvindex are not statistically significant because its p-value is larger than 0.05. Therefore, we reject the null hypothesis and conclude that at least one group having a significant difference on mean.

6.3.4 Multivariate Analysis

We examine the relationship between multiple independent variables and the dependent variable using multiple linear regression and ANOVA test. After that, we show the distribution of the model residuals using histograms.

```
Analysis of Variance Table

Response: precip
               Df Sum Sq Mean Sq F value    Pr(>F)    
description     8  39139  4892.4 37.0930 < 2.2e-16 ***
uvindex         9    825    91.6  0.6948 0.7140506    
tempmax         1   1228  1227.9  9.3094 0.0023419 ** 
tempmin         1   4485  4484.9 34.0034 7.490e-09 ***
temp            1   5891  5890.6 44.6615 3.946e-11 ***
feelslikemax    1    577   577.1  4.3756 0.0367174 *  
feelslikemin    1   1364  1363.8 10.3399 0.0013450 ** 
feelslike       1    117   116.9  0.8861 0.3467760    
dew             1   1795  1795.3 13.6115 0.0002373 ***
humidity        1   9948  9948.1 75.4242 < 2.2e-16 ***
precipcover     1      7     6.6  0.0501 0.8229131    
cloudcover      1    475   475.4  3.6043 0.0579260 .  
visibility      1   1245  1244.6  9.4359 0.0021871 ** 
solarradiation  1    834   833.6  6.3203 0.0120976 *  
solarenergy     1    130   130.2  0.9871 0.3207109    
Residuals     969 127807   131.9                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: ANOVA test Result for Multiple Linear Regression

H_0 : All group means are equal.

H_1 : At least one group mean is different

The histogram above skewed to the right, and we cannot conclude the normality with enough confidence. Thus, we plot the residues on the normal Q-Q plot. From the normal Q-Q plot, we can observe that most residuals lie in a straight line which indicates that it follows a normal distribution.

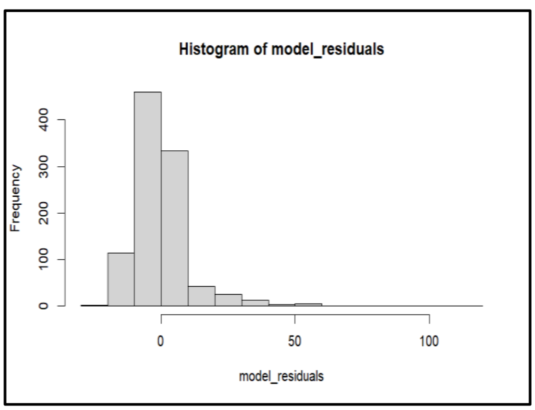


Figure 6:Histogram of Model Residuals

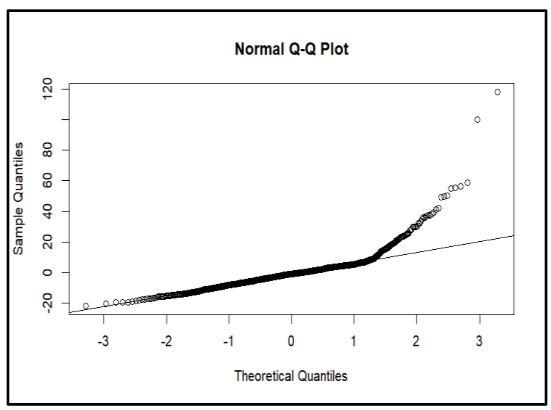


Figure 6: Normal Q-Q Plot

7.0 Ethical Consideration

Below are some of the ethical considerations need to be taken into account for this project:

1. Transparency in Data Processing will be prioritized by maintaining transparency in data processing steps, clearly documenting and communicating the methods used for data cleaning, imputation, and feature engineering.
2. Equitable Representation in Model Training will be highlighted to ensure equitable representation of diverse meteorological conditions within Kuala Lumpur during the training phase of the machine learning model. Prevention of over-reliance on specific weather patterns shall be considered to hinder biased predictions produced, especially in cases of extreme events.

8.0 Impact of the Project to the Society

The implementation of accurate rainfall predictions in Kuala Lumpur using advanced machine learning models would benefit various sectors, significantly enhancing the region's resilience and sustainability.

Firstly, in the agricultural sector, precise forecasts empower farmers with invaluable information for optimal crop planning and irrigation management. Aside from improved agricultural productivity, it also minimizes resource wastage and provides a more sustainable approach to farming amidst climate-induced uncertainties.

Additionally, the project facilitates more effective water resource management, enabling local authorities to plan and allocate resources based on anticipated rainfall patterns. This contributes to sustainable water usage for both urban and agricultural needs, addressing critical water scarcity challenges.

Timely and accurate rainfall predictions also play a vital role in flood risk mitigation, and rainfall would be one of the crucial indicators in implementing preventive measures and reducing the impact of floods.

The project also contributes towards urban planning and infrastructure design, where reliable forecasts inform the development of resilient systems, including drainage and flood control. Moreover, an improved forecasting would be beneficial in developing an efficient disaster response plan and may potentially reduce economic costs associated with emergency services, such as flood evacuation services.

Beyond these practical benefits, the project fosters community empowerment and resilience by providing accurate weather information. People can make informed decisions, take proactive measures, and be well-prepared for climate-related challenges.

Most importantly, the project aligns with global efforts toward climate action, addressing a critical aspect of climate change adaptation. Accurate rainfall prediction supports local and national climate resilience strategies, contributing to broader initiatives aimed at mitigating the impact of climate change. Furthermore, businesses may also take advantage from this project outcome, especially those reliant on weather conditions. Industries such as tourism, construction, and logistics can optimize operations based on reliable weather forecasts and make data-driven decisions.

In summary, the implementation of accurate rainfall predictions in Kuala Lumpur is capable of offering a transformative approach that positively impacts agriculture, water management, disaster resilience, urban planning, climate action and business decision-making.

References

[1] Aderyani, F. R., Mousavi, S. J., & Jafari, F. (2022). Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN. *Journal of Hydrology, 614*(Part A), 128463. doi:https://doi.org/10.1016/j.jhydrol.2022.128463

[2] Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, 100204.doi: https://doi.org/10.1016/j.mlwa.2021.100204

[3] Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00545-4

[4] Mohammed, M., Kolapalli, R., Golla, N., & Maturi, S. S. (2020). Prediction Of Rainfall Using Machine Learning Techniques. *International Journal Of Scientific & Technology Research, 9*(01), 3236-3240.Retrieved from https://www.semanticscholar.org/paper/Prediction-Of-Rainfall-Using-Machine-Learning-MohammedKolapalli/9bf0a85ff6323559663cb45c89ae1e51684401dc

[5] Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal, 12*(2), 1651-1663. doi:https://doi.org/10.1016/j.asej.2020.09.011

[6] Sulaiman, N. A., Shaharudin, S. M., Ismail, S., Zainuddin, N. H., Tan, M. L., & Abd Jalil, Y. (2022). Predictive modelling of statistical downscaling based on hybrid machine learning model for daily rainfall in east-coast Peninsular Malaysia. *Symmetry*, *14*(5), 927. doi:https://doi.org/10.3390/sym14050927

Appendix

Rainfall

CHEN XIN WEI 17202767

2023-12-09

Contents

0.1 Get data

#Data Exploration

```
names(Rainfall)
```

```
##  [1] "name"             "datetime"         "tempmax"          "tempmin"
##  [5] "temp"             "feelslikemax"     "feelslikemin"     "feelslike"
##  [9] "dew"              "humidity"         "precip"           "precipprob"
## [13] "precipcover"      "preciptype"       "snow"             "snowdepth"
## [17] "windgust"         "windspeed"        "winddir"          "sealevelpressure"
## [21] "cloudcover"       "visibility"       "solarradiation"   "solarenergy"
## [25] "uvindex"          "severerisk"       "sunrise"          "sunset"
## [29] "moonphase"        "conditions"       "description"      "icon"
## [33] "stations"
```

```
class(Rainfall)
```

```
## [1] "data.frame"
```

```
str(Rainfall)
```

```
## 'data.frame':    1000 obs. of  33 variables:
##  $ name            : chr  "Kuala Lumpur,Malaysia" "Kuala Lumpur,Malaysia" "Kuala Lumpur,Malaysia" "Ku
##  $ datetime        : chr  "2021-02-01" "2021-02-02" "2021-02-03" "2021-02-04" ...
##  $ tempmax         : num  33.1 34 33.8 33.1 33.9 34.1 34.8 33.8 32.9 32.9 ...
##  $ tempmin         : num  24.1 25.8 24.9 24.1 22.9 24.1 24.9 25 24.9 24.8 ...
```

```
##  $ temp            : num  27.5 28.8 28 28.3 27.9 28.5 28.6 27.7 28.6 28.4 ...
##  $ feelslikemax    : num  36.2 37.3 38.6 36.3 36 38.2 40.4 39 39.1 37.2 ...
##  $ feelslikemin    : num  24.1 25.8 24.9 24.1 22.9 24.1 24.9 25 24.9 24.8 ...
##  $ feelslike       : num  28.8 31.3 30.5 29.7 29.3 30.7 31 29.6 31.8 31.6 ...
##  $ dew             : num  22.4 22.2 23.2 21.7 20.6 22.5 22.9 23.4 23.8 24 ...
##  $ humidity        : num  76.4 69.1 76.5 69.5 66.5 72.2 72.8 78.6 76.8 77.9 ...
##  $ precip          : num  47.739 0 0.707 0 0 ...
##  $ precipprob      : int  100 0 100 0 0 100 0 100 0 100 ...
##  $ precipcover     : num  4.17 0 4.17 0 0 ...
##  $ preciptype      : chr  "rain" "" "rain" "" ...
##  $ snow            : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ snowdepth       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ windgust        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ windspeed       : num  19 12.5 15.7 8.7 14.3 14.6 17.7 15.7 17.3 13.9 ...
##  $ winddir         : num  67.3 5.7 294 40.5 307.2 ...
##  $ sealevelpressure: num  1012 1012 1011 1011 1011 ...
##  $ cloudcover      : num  41.2 37.9 39.6 37.3 37.2 38.6 37.6 40.1 38.2 34.7 ...
##  $ visibility      : num  9.5 10.2 10 9.5 10.2 10.1 9.9 9.9 9.5 9.7 ...
##  $ solarradiation  : num  258 264 266 293 298 ...
##  $ solarenergy     : num  22.1 22.7 22.9 25.2 25.6 24.9 25.8 23.9 24 19.3 ...
##  $ uvindex         : int  9 9 9 10 10 10 10 10 10 7 ...
##  $ severerisk      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ sunrise         : chr  "2021-02-01T07:27:12" "2021-02-02T07:27:16" "2021-02-03T07:27:19" "2021-02-
##  $ sunset          : chr  "2021-02-01T19:26:28" "2021-02-02T19:26:39" "2021-02-03T19:26:50" "2021-02-
##  $ moonphase       : num  0.62 0.66 0.69 0.73 0.75 0.8 0.83 0.87 0.91 0.94 ...
##  $ conditions      : chr  "Rain, Partially cloudy" "Partially cloudy" "Rain, Partially cloudy" "Parti
##  $ description     : chr  "Partly cloudy throughout the day with early morning rain." "Partly cloudy
##  $ icon            : chr  "rain" "partly-cloudy-day" "rain" "partly-cloudy-day" ...
##  $ stations        : chr  "48647099999,48650099999,WMSA,WMKK" "48647099999,48650099999,WMSA,WMKK" "48
```

```r
dim(Rainfall)
```

```
## [1] 1000   33
```

```r
summary(Rainfall)
```

```
##      name             datetime            tempmax         tempmin
##  Length:1000        Length:1000        Min.   :25.70   Min.   :21.4
##  Class :character   Class :character   1st Qu.:31.90   1st Qu.:24.1
##  Mode  :character   Mode  :character   Median :32.90   Median :24.9
##                                        Mean   :32.59   Mean   :24.9
##                                        3rd Qu.:33.80   3rd Qu.:25.7
##                                        Max.   :35.80   Max.   :27.9
##
##       temp        feelslikemax    feelslikemin     feelslike
##  Min.   :24.30   Min.   :25.70   Min.   :21.40   Min.   :24.30
##  1st Qu.:27.30   1st Qu.:36.80   1st Qu.:24.10   1st Qu.:29.30
##  Median :28.00   Median :38.10   Median :24.90   Median :30.60
##  Mean   :28.01   Mean   :38.14   Mean   :25.07   Mean   :30.72
##  3rd Qu.:28.80   3rd Qu.:39.60   3rd Qu.:25.70   3rd Qu.:32.00
##  Max.   :31.10   Max.   :45.30   Max.   :32.20   Max.   :36.80
##
##       dew           humidity         precip          precipprob
```

```
##  Min.   :19.90   Min.   :61.70   Min.   :  0.0000   Min.   :  0.0
##  1st Qu.:23.20   1st Qu.:74.70   1st Qu.:  0.0000   1st Qu.:  0.0
##  Median :23.70   Median :78.60   Median :  0.5535   Median :100.0
##  Mean   :23.62   Mean   :78.47   Mean   :  6.5294   Mean   : 65.4
##  3rd Qu.:24.10   3rd Qu.:82.20   3rd Qu.:  6.1513   3rd Qu.:100.0
##  Max.   :25.50   Max.   :95.70   Max.   :168.0730   Max.   :100.0
##
##   precipcover     preciptype            snow       snowdepth
##  Min.   : 0.000   Length:1000        Min.   :0     Min.   :0
##  1st Qu.: 0.000   Class :character   1st Qu.:0     1st Qu.:0
##  Median : 4.170   Mode  :character   Median :0     Median :0
##  Mean   : 4.134                      Mean   :0     Mean   :0
##  3rd Qu.: 4.170                      3rd Qu.:0     3rd Qu.:0
##  Max.   :83.330                      Max.   :0     Max.   :0
##                                      NA's   :343   NA's   :343
##     windgust       windspeed        winddir      sealevelpressure
##  Min.   : 3.20   Min.   : 5.40   Min.   :  3.1   Min.   :1006
##  1st Qu.: 8.70   1st Qu.:13.50   1st Qu.:171.3   1st Qu.:1009
##  Median :11.20   Median :15.50   Median :263.3   Median :1010
##  Mean   :15.06   Mean   :15.69   Mean   :237.3   Mean   :1010
##  3rd Qu.:15.10   3rd Qu.:17.50   3rd Qu.:305.6   3rd Qu.:1010
##  Max.   :59.40   Max.   :38.00   Max.   :358.8   Max.   :1015
##  NA's   :310                                     NA's   :1
##    cloudcover     visibility    solarradiation   solarenergy
##  Min.   :25.8   Min.   : 5.70   Min.   : 22.6   Min.   : 1.80
##  1st Qu.:38.2   1st Qu.: 9.30   1st Qu.:178.1   1st Qu.:15.40
##  Median :39.8   Median : 9.70   Median :225.4   Median :19.40
##  Mean   :42.5   Mean   : 9.57   Mean   :214.0   Mean   :18.47
##  3rd Qu.:47.4   3rd Qu.:10.00   3rd Qu.:256.8   3rd Qu.:22.20
##  Max.   :58.1   Max.   :10.50   Max.   :317.8   Max.   :27.50
##
##     uvindex         severerisk     sunrise             sunset
##  Min.   : 1.000   Min.   :10.0   Length:1000        Length:1000
##  1st Qu.: 7.000   1st Qu.:30.0   Class :character   Class :character
##  Median : 8.000   Median :30.0   Mode  :character   Mode  :character
##  Mean   : 7.887   Mean   :37.6
##  3rd Qu.: 9.000   3rd Qu.:60.0
##  Max.   :10.000   Max.   :75.0
##                   NA's   :343
##    moonphase       conditions         description            icon
##  Min.   :0.0000   Length:1000        Length:1000        Length:1000
##  1st Qu.:0.2500   Class :character   Class :character   Class :character
##  Median :0.4800   Mode  :character   Mode  :character   Mode  :character
##  Mean   :0.4839
##  3rd Qu.:0.7500
##  Max.   :0.9800
##
##    stations
##  Length:1000
##  Class :character
##  Mode  :character
##
##
##
```

```
##
```

```r
head(Rainfall)
```

```
##                       name   datetime tempmax tempmin temp feelslikemax
## 1 Kuala Lumpur,Malaysia 2021-02-01    33.1    24.1 27.5         36.2
## 2 Kuala Lumpur,Malaysia 2021-02-02    34.0    25.8 28.8         37.3
## 3 Kuala Lumpur,Malaysia 2021-02-03    33.8    24.9 28.0         38.6
## 4 Kuala Lumpur,Malaysia 2021-02-04    33.1    24.1 28.3         36.3
## 5 Kuala Lumpur,Malaysia 2021-02-05    33.9    22.9 27.9         36.0
## 6 Kuala Lumpur,Malaysia 2021-02-06    34.1    24.1 28.5         38.2
##   feelslikemin feelslike  dew humidity precip precipprob precipcover preciptype
## 1         24.1      28.8 22.4     76.4 47.739        100        4.17       rain
## 2         25.8      31.3 22.2     69.1  0.000          0        0.00
## 3         24.9      30.5 23.2     76.5  0.707        100        4.17       rain
## 4         24.1      29.7 21.7     69.5  0.000          0        0.00
## 5         22.9      29.3 20.6     66.5  0.000          0        0.00
## 6         24.1      30.7 22.5     72.2  0.354        100        4.17       rain
##   snow snowdepth windgust windspeed winddir sealevelpressure cloudcover
## 1   NA        NA       NA      19.0    67.3           1011.6       41.2
## 2   NA        NA       NA      12.5     5.7           1011.6       37.9
## 3   NA        NA       NA      15.7   294.0           1011.1       39.6
## 4   NA        NA       NA       8.7    40.5           1011.2       37.3
## 5   NA        NA       NA      14.3   307.2           1010.9       37.2
## 6   NA        NA       NA      14.6   111.2           1009.8       38.6
##   visibility solarradiation solarenergy uvindex severerisk             sunrise
## 1        9.5          257.9        22.1       9         NA 2021-02-01T07:27:12
## 2       10.2          263.6        22.7       9         NA 2021-02-02T07:27:16
## 3       10.0          266.3        22.9       9         NA 2021-02-03T07:27:19
## 4        9.5          293.2        25.2      10         NA 2021-02-04T07:27:21
## 5       10.2          297.8        25.6      10         NA 2021-02-05T07:27:22
## 6       10.1          288.5        24.9      10         NA 2021-02-06T07:27:23
##                sunset moonphase             conditions
## 1 2021-02-01T19:26:28      0.62 Rain, Partially cloudy
## 2 2021-02-02T19:26:39      0.66       Partially cloudy
## 3 2021-02-03T19:26:50      0.69 Rain, Partially cloudy
## 4 2021-02-04T19:27:00      0.73       Partially cloudy
## 5 2021-02-05T19:27:09      0.75       Partially cloudy
## 6 2021-02-06T19:27:18      0.80 Rain, Partially cloudy
##                                                  description              icon
## 1  Partly cloudy throughout the day with early morning rain.              rain
## 2                         Partly cloudy throughout the day. partly-cloudy-day
## 3 Partly cloudy throughout the day with late afternoon rain.              rain
## 4                         Partly cloudy throughout the day. partly-cloudy-day
## 5                         Partly cloudy throughout the day. partly-cloudy-day
## 6 Partly cloudy throughout the day with late afternoon rain.              rain
##                         stations
## 1 48647099999,48650099999,WMSA,WMKK
## 2 48647099999,48650099999,WMSA,WMKK
## 3 48647099999,48650099999,WMSA,WMKK
## 4 48647099999,48650099999,WMSA,WMKK
## 5 48647099999,48650099999,WMSA,WMKK
## 6 48647099999,48650099999,WMSA,WMKK
```

```r
tail(Rainfall)
```

```
##                        name   datetime tempmax tempmin temp feelslikemax
## 995  Kuala Lumpur,Malaysia 2023-10-23    33.7    24.0 27.0         39.4
## 996  Kuala Lumpur,Malaysia 2023-10-24    31.9    24.8 26.5         37.5
## 997  Kuala Lumpur,Malaysia 2023-10-25    32.6    24.8 27.5         38.6
## 998  Kuala Lumpur,Malaysia 2023-10-26    33.0    24.8 26.7         39.3
## 999  Kuala Lumpur,Malaysia 2023-10-27    33.1    25.0 27.4         40.3
## 1000 Kuala Lumpur,Malaysia 2023-10-28    33.4    25.0 26.8         39.8
##      feelslikemin feelslike  dew humidity precip precipprob precipcover
## 995          24.0      28.8 23.3     82.0  4.884        100        4.17
## 996          24.8      27.6 23.9     86.3 23.056        100        8.33
## 997          24.8      29.6 24.0     82.3  2.232        100        4.17
## 998          24.8      28.0 24.3     87.5 16.000        100       50.00
## 999          25.0      29.2 24.0     83.3  5.700        100       54.17
## 1000         25.0      28.4 24.1     86.4 23.500        100       83.33
##      preciptype snow snowdepth windgust windspeed winddir sealevelpressure
## 995        rain    0         0      7.6      19.9   310.6           1012.0
## 996        rain    0         0      9.0      16.6   311.5           1011.9
## 997        rain    0         0      8.3      19.6   308.3           1012.3
## 998        rain    0         0      8.3      17.1   255.2           1011.7
## 999        rain    0         0     11.9      18.5   302.1           1010.8
## 1000       rain    0         0     38.9      18.1    77.3           1010.6
##      cloudcover visibility solarradiation solarenergy uvindex severerisk
## 995        50.0        9.4          256.0        22.2      10         60
## 996        48.0        9.3          234.2        20.1      10         30
## 997        52.7        9.6          276.7        24.0      10         30
## 998        30.2        9.4          276.1        23.7       9         30
## 999        25.8        7.8          221.9        19.2      10         30
## 1000       28.1        7.3          230.3        19.9       9         30
##                  sunrise              sunset moonphase            conditions
## 995  2023-10-23T06:56:44 2023-10-23T18:58:30      0.30 Rain, Partially cloudy
## 996  2023-10-24T06:56:40 2023-10-24T18:58:17      0.33 Rain, Partially cloudy
## 997  2023-10-25T06:56:37 2023-10-25T18:58:05      0.37 Rain, Partially cloudy
## 998  2023-10-26T06:56:34 2023-10-26T18:57:53      0.40 Rain, Partially cloudy
## 999  2023-10-27T06:56:33 2023-10-27T18:57:43      0.44 Rain, Partially cloudy
## 1000 2023-10-28T06:56:32 2023-10-28T18:57:33      0.48 Rain, Partially cloudy
##                                                                   description
## 995                    Partly cloudy throughout the day with early morning rain.
## 996                                 Partly cloudy throughout the day with rain.
## 997                          Partly cloudy throughout the day with morning rain.
## 998                                 Partly cloudy throughout the day with rain.
## 999  Partly cloudy throughout the day with a chance of rain throughout the day.
## 1000 Partly cloudy throughout the day with a chance of rain throughout the day.
##      icon                      stations
## 995  rain 48647099999,48650099999,WMSA,WMKK
## 996  rain 48647099999,48650099999,WMSA,WMKK
## 997  rain 48647099999,48650099999,WMSA,WMKK
## 998  rain 48647099999,48650099999,WMSA,WMKK
## 999  rain                         WMSA,WMKK
## 1000 rain                         WMSA,WMKK
```

#First Variable Drop

```r
Rainfall <- subset(Rainfall,
            select = -c(name,datetime,precipprob,preciptype,snow,
                        snowdepth,sunrise,sunset,conditions,stations))
str(Rainfall)
```

```
## 'data.frame':    1000 obs. of  23 variables:
##  $ tempmax         : num  33.1 34 33.8 33.1 33.9 34.1 34.8 33.8 32.9 32.9 ...
##  $ tempmin         : num  24.1 25.8 24.9 24.1 22.9 24.1 24.9 25 24.9 24.8 ...
##  $ temp            : num  27.5 28.8 28 28.3 27.9 28.5 28.6 27.7 28.6 28.4 ...
##  $ feelslikemax    : num  36.2 37.3 38.6 36.3 36 38.2 40.4 39 39.1 37.2 ...
##  $ feelslikemin    : num  24.1 25.8 24.9 24.1 22.9 24.1 24.9 25 24.9 24.8 ...
##  $ feelslike       : num  28.8 31.3 30.5 29.7 29.3 30.7 31 29.6 31.8 31.6 ...
##  $ dew             : num  22.4 22.2 23.2 21.7 20.6 22.5 22.9 23.4 23.8 24 ...
##  $ humidity        : num  76.4 69.1 76.5 69.5 66.5 72.2 72.8 78.6 76.8 77.9 ...
##  $ precip          : num  47.739 0 0.707 0 0 ...
##  $ precipcover     : num  4.17 0 4.17 0 0 ...
##  $ windgust        : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ windspeed       : num  19 12.5 15.7 8.7 14.3 14.6 17.7 15.7 17.3 13.9 ...
##  $ winddir         : num  67.3 5.7 294 40.5 307.2 ...
##  $ sealevelpressure: num  1012 1012 1011 1011 1011 ...
##  $ cloudcover      : num  41.2 37.9 39.6 37.3 37.2 38.6 37.6 40.1 38.2 34.7 ...
##  $ visibility      : num  9.5 10.2 10 9.5 10.2 10.1 9.9 9.9 9.5 9.7 ...
##  $ solarradiation  : num  258 264 266 293 298 ...
##  $ solarenergy     : num  22.1 22.7 22.9 25.2 25.6 24.9 25.8 23.9 24 19.3 ...
##  $ uvindex         : int  9 9 9 10 10 10 10 10 10 7 ...
##  $ severerisk      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ moonphase       : num  0.62 0.66 0.69 0.73 0.75 0.8 0.83 0.87 0.91 0.94 ...
##  $ description     : chr  "Partly cloudy throughout the day with early morning rain." "Partly cloudy
##  $ icon            : chr  "rain" "partly-cloudy-day" "rain" "partly-cloudy-day" ...
```

#Transform categorical variable

```r
Rainfall$description<-as.factor(Rainfall$description)
Rainfall$icon<-as.factor(Rainfall$icon)
Rainfall$uvindex<-as.factor(Rainfall$uvindex)
```

##Checking relationship between cat var with dependent var

```r
corr<-aov(precip ~ description * uvindex * icon, data=Rainfall)
summary(corr)
```

```
##                      Df Sum Sq Mean Sq F value Pr(>F)
## description           8  39139    4892  30.606 <2e-16 ***
## uvindex               9    825      92   0.573  0.820
## description:uvindex  51   7079     139   0.868  0.732
## Residuals           931 148822     160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#correlation matric

```r
Rainfall_sub <- subset(Rainfall, select = -c(description,icon,uvindex))
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.3.2
```

```
## corrplot 0.92 loaded
```

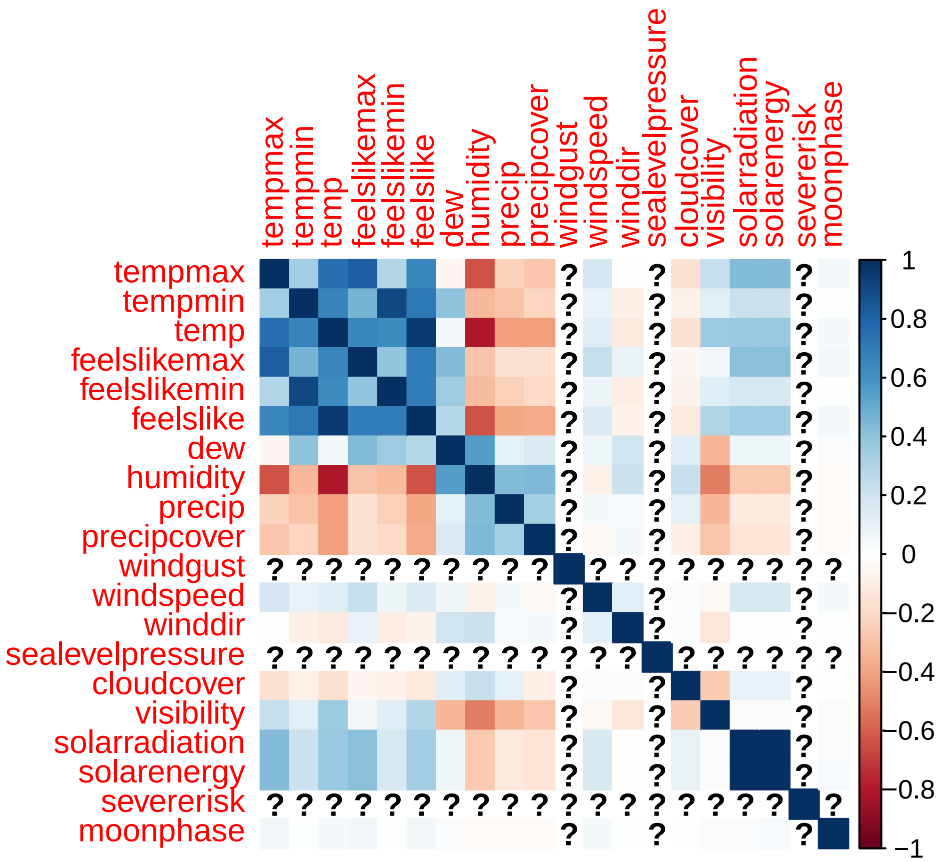
```r
cr_Rainfall <- cor(Rainfall_sub)
cr_Rainfall
```

```
##                       tempmax      tempmin        temp feelslikemax
## tempmax           1.000000000  0.346286275  0.75697724   0.82183941
## tempmin           0.346286275  1.000000000  0.66498921   0.46399103
## temp              0.756977238  0.664989209  1.00000000   0.65925566
## feelslikemax      0.821839415  0.463991031  0.65925566   1.00000000
## feelslikemin      0.298867339  0.909092422  0.62148532   0.39844155
## feelslike         0.655585169  0.716528872  0.95234958   0.69867355
## dew              -0.049046346  0.409861903  0.04211583   0.43098257
## humidity         -0.635125753 -0.323443217 -0.80675532  -0.28192083
## precip           -0.228235231 -0.287535073 -0.41880522  -0.15148459
## precipcover      -0.276169757 -0.216538872 -0.41953149  -0.15795598
## windgust                   NA           NA          NA           NA
## windspeed         0.185626044  0.098122898  0.13824272   0.23592413
## winddir          -0.007335355 -0.083734130 -0.11818292   0.09684709
## sealevelpressure           NA           NA          NA           NA
## cloudcover       -0.151926306 -0.073853416 -0.16376114  -0.04422532
## visibility        0.232677360  0.125655224  0.36044288   0.05390195
## solarradiation    0.438493948  0.214263651  0.37407025   0.41577111
## solarenergy       0.439040783  0.215774508  0.37602754   0.41571567
## severerisk                 NA           NA          NA           NA
## moonphase         0.051458584  0.005763154  0.04652635   0.05795330
##                  feelslikemin   feelslike         dew    humidity      precip
## tempmax           0.298867339  0.65558517 -0.04904635 -0.63512575 -0.22823523
## tempmin           0.909092422  0.71652887  0.40986190 -0.32344322 -0.28753507
## temp              0.621485317  0.95234958  0.04211583 -0.80675532 -0.41880522
## feelslikemax      0.398441552  0.69867355  0.43098257 -0.28192083 -0.15148459
## feelslikemin      1.000000000  0.69017903  0.35123283 -0.31942190 -0.23785457
## feelslike         0.690179031  1.00000000  0.28192648 -0.63424905 -0.38041634
## dew               0.351232826  0.28192648  1.00000000  0.55043979  0.10831756
## humidity         -0.319421901 -0.63424905  0.55043979  1.00000000  0.43586047
## precip           -0.237854565 -0.38041634  0.10831756  0.43586047  1.00000000
## precipcover      -0.199202915 -0.37050155  0.15098855  0.44660758  0.33492667
## windgust                   NA          NA          NA          NA          NA
## windspeed         0.077500198  0.14584869  0.06417030 -0.07411025  0.04802204
## winddir          -0.096307935 -0.06845136  0.19146693  0.21375965  0.03359117
## sealevelpressure           NA          NA          NA          NA          NA
## cloudcover       -0.069321499 -0.11543448  0.13981211  0.22175313  0.11285197
## visibility        0.139598477  0.29322085 -0.33233999 -0.51885448 -0.33523942
## solarradiation    0.174907437  0.34224519  0.06210029 -0.26352374 -0.11037574
## solarenergy       0.176218568  0.34407816  0.06136088 -0.26574257 -0.11273581
## severerisk                 NA          NA          NA          NA          NA
## moonphase         0.005589589  0.05162397  0.02387654 -0.02616104 -0.02378210
```

```
##                  precipcover windgust   windspeed       winddir
## tempmax          -0.27616976       NA  0.18562604 -0.0073353552
## tempmin          -0.21653887       NA  0.09812290 -0.0837341302
## temp             -0.41953149       NA  0.13824272 -0.1181829186
## feelslikemax     -0.15795598       NA  0.23592413  0.0968470926
## feelslikemin     -0.19920292       NA  0.07750020 -0.0963079346
## feelslike        -0.37050155       NA  0.14584869 -0.0684513597
## dew               0.15098855       NA  0.06417030  0.1914669252
## humidity          0.44660758       NA -0.07411025  0.2137596532
## precip            0.33492667       NA  0.04802204  0.0335911681
## precipcover       1.00000000       NA -0.03990282  0.0413750578
## windgust                  NA        1          NA            NA
## windspeed        -0.03990282       NA  1.00000000  0.1295410134
## winddir           0.04137506       NA  0.12954101  1.0000000000
## sealevelpressure          NA       NA          NA            NA
## cloudcover       -0.08821699       NA  0.02145366  0.0135095683
## visibility       -0.27099911       NA -0.03972945 -0.1313638011
## solarradiation   -0.14227183       NA  0.16195969 -0.0016971329
## solarenergy      -0.14449900       NA  0.16106639 -0.0026558909
## severerisk                NA       NA          NA            NA
## moonphase        -0.02746689       NA  0.04953837  0.0004913426
##                  sealevelpressure   cloudcover  visibility solarradiation
## tempmax                        NA -0.151926306  0.23267736    0.438493948
## tempmin                        NA -0.073853416  0.12565522    0.214263651
## temp                           NA -0.163761137  0.36044288    0.374070249
## feelslikemax                   NA -0.044225324  0.05390195    0.415771111
## feelslikemin                   NA -0.069321499  0.13959848    0.174907437
## feelslike                      NA -0.115434483  0.29322085    0.342245187
## dew                            NA  0.139812114 -0.33233999    0.062100290
## humidity                       NA  0.221753131 -0.51885448   -0.263523736
## precip                         NA  0.112851966 -0.33523942   -0.110375742
## precipcover                    NA -0.088216993 -0.27099911   -0.142271826
## windgust                       NA           NA          NA             NA
## windspeed                      NA  0.021453660 -0.03972945    0.161959688
## winddir                        NA  0.013509568 -0.13136380   -0.001697133
## sealevelpressure                1           NA          NA             NA
## cloudcover                     NA  1.000000000 -0.25781629    0.093954154
## visibility                     NA -0.257816286  1.00000000    0.018634019
## solarradiation                 NA  0.093954154  0.01863402    1.000000000
## solarenergy                    NA  0.094312815  0.01947126    0.999699051
## severerisk                     NA           NA          NA             NA
## moonphase                      NA  0.004187397  0.02269174    0.029699572
##                   solarenergy severerisk     moonphase
## tempmax           0.439040783         NA  0.0514585839
## tempmin           0.215774508         NA  0.0057631535
## temp              0.376027537         NA  0.0465263499
## feelslikemax      0.415715671         NA  0.0579533047
## feelslikemin      0.176218568         NA  0.0055895892
## feelslike         0.344078164         NA  0.0516239725
## dew               0.061360875         NA  0.0238765359
## humidity         -0.265742572         NA -0.0261610447
## precip           -0.112735807         NA -0.0237821045
## precipcover      -0.144499000         NA -0.0274668863
## windgust                   NA         NA            NA
```

```
## windspeed          0.161066392          NA  0.0495383667
## winddir           -0.002655891          NA  0.0004913426
## sealevelpressure            NA          NA            NA
## cloudcover         0.094312815          NA  0.0041873973
## visibility         0.019471256          NA  0.0226917435
## solarradiation     0.999699051          NA  0.0296995721
## solarenergy        1.000000000          NA  0.0313434266
## severerisk                  NA           1            NA
## moonphase          0.031343427          NA  1.0000000000
```

```r
corrplot(cr_Rainfall,method="color")
```



0.2 Second Variable Drop

```r
Rainfall <- subset(Rainfall,
            select = -c(windgust, windspeed, winddir, sealevelpressure,
                        moonphase, severerisk, icon))
```

0.3 Check missing value

```r
sum(is.na(Rainfall))
```

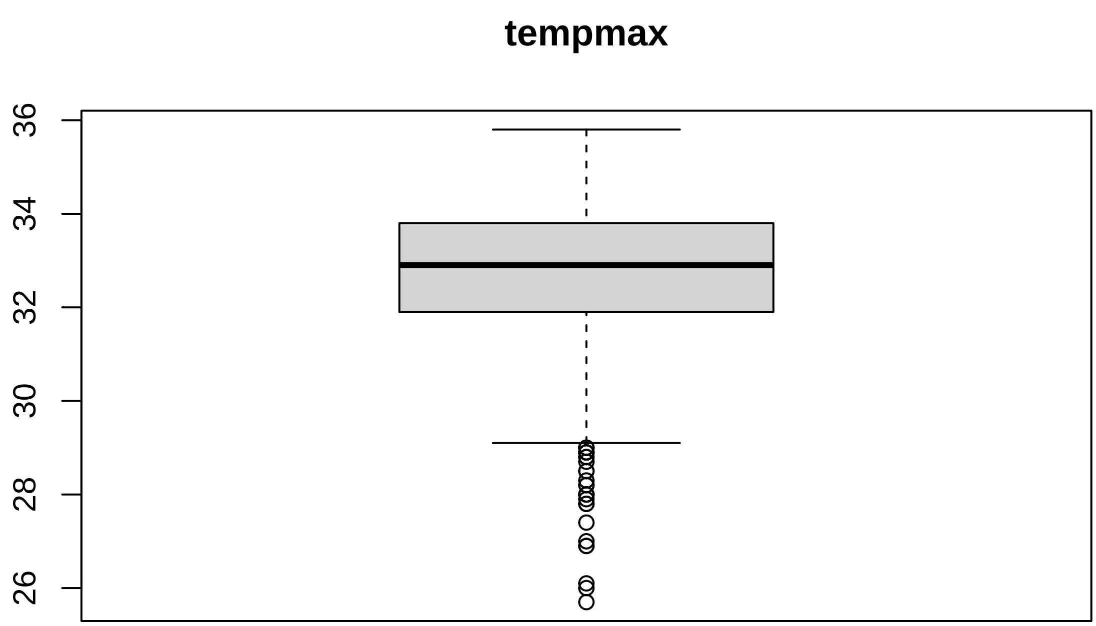
```
## [1] 0
```

```r
colSums(is.na(Rainfall))
```

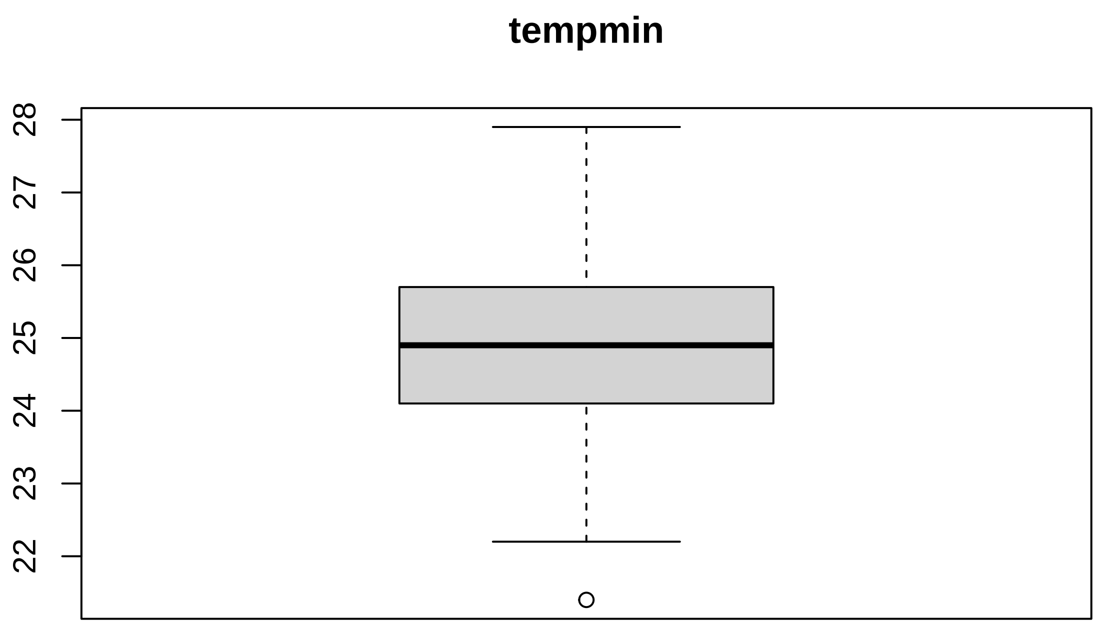
```
##         tempmax         tempmin            temp     feelslikemax    feelslikemin
##               0               0               0                0               0
##       feelslike             dew        humidity          precip     precipcover
##               0               0               0                0               0
##      cloudcover      visibility solarradiation     solarenergy         uvindex
##               0               0               0                0               0
##     description
##               0
```

0.4 Check outliers

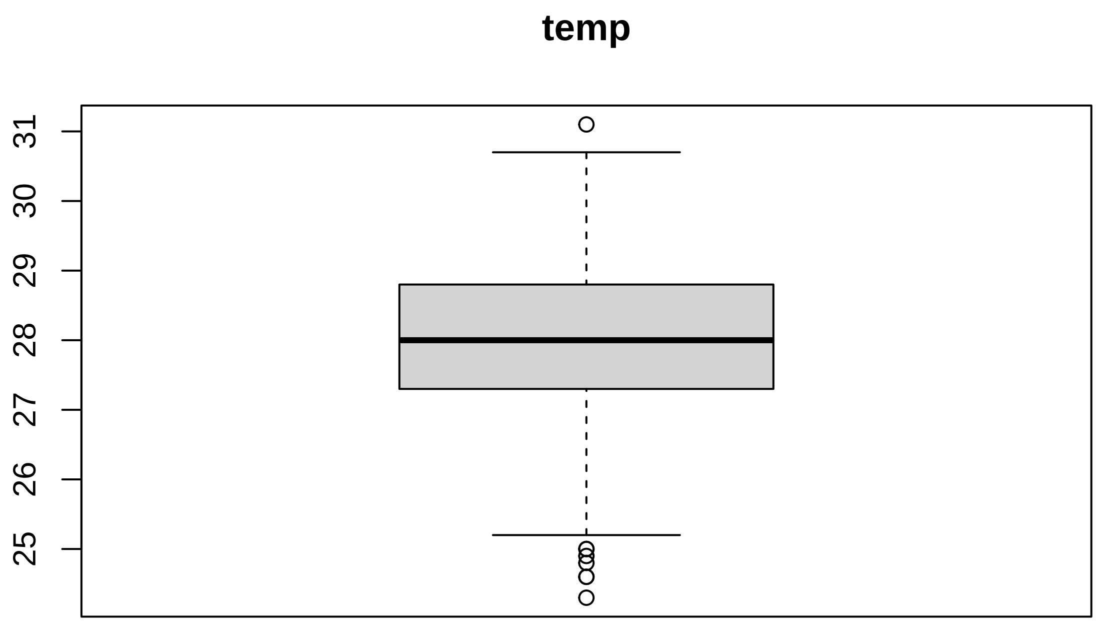
```r
boxplot(Rainfall$tempmax)
title("tempmax")
```



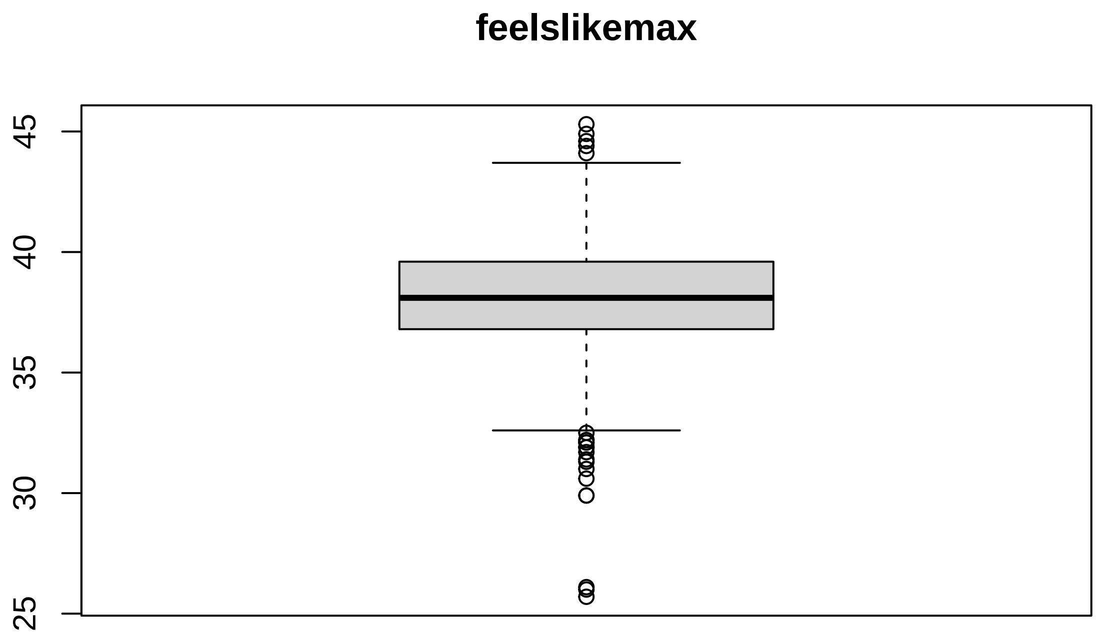
```r
boxplot(Rainfall$tempmin)
title("tempmin")
```



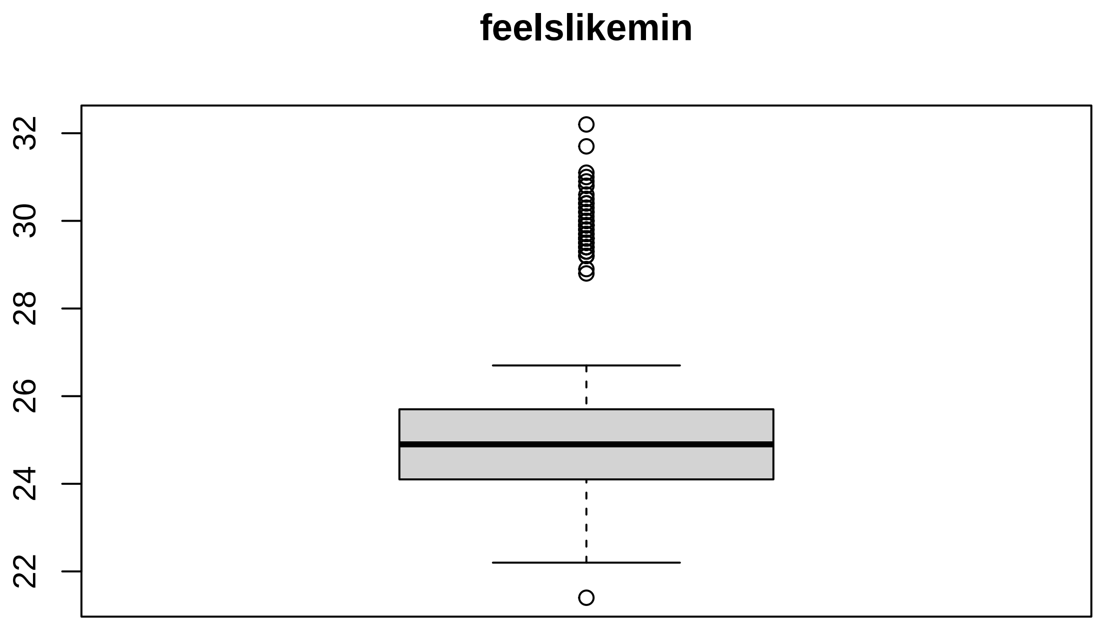

```r
boxplot(Rainfall$temp)
title("temp")
```



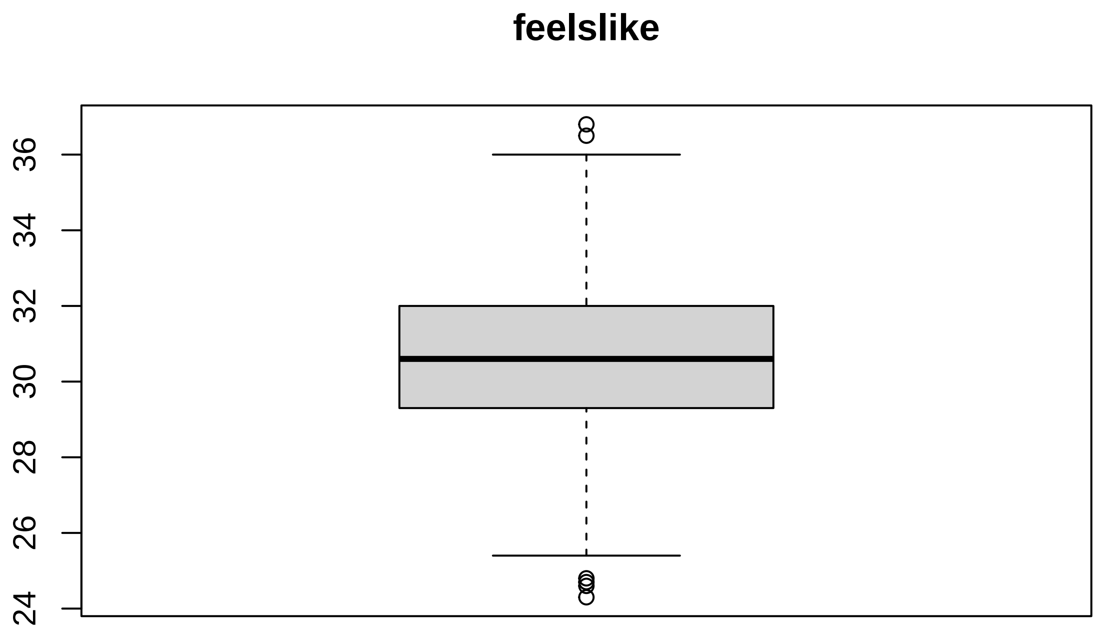

```r
boxplot(Rainfall$feelslikemax)
title("feelslikemax")
```



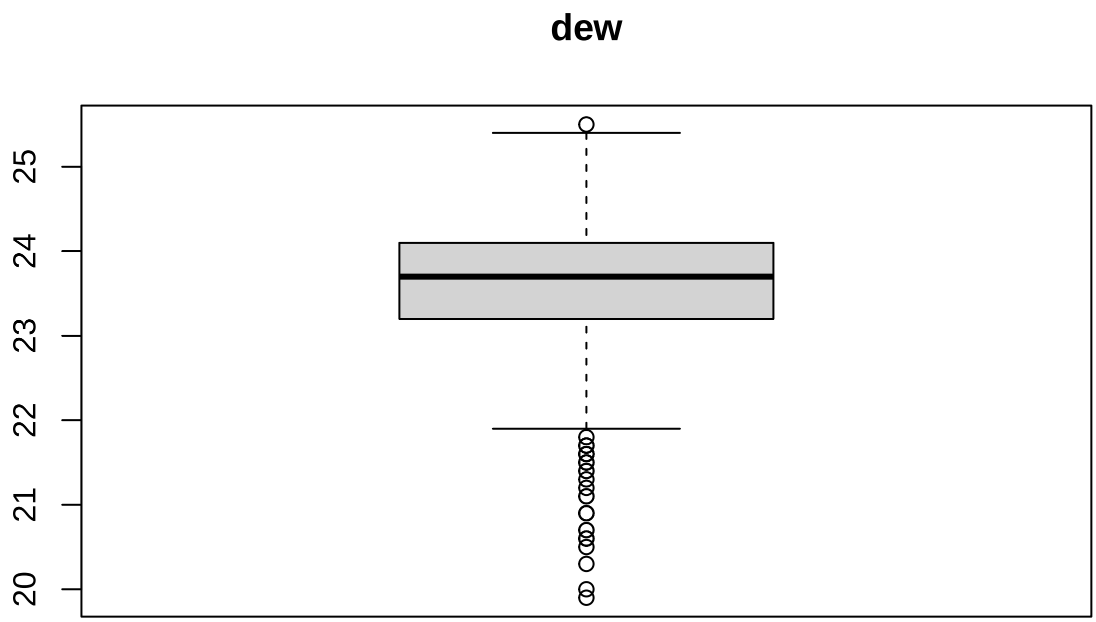

```
boxplot(Rainfall$feelslikemin)
title("feelslikemin")
```



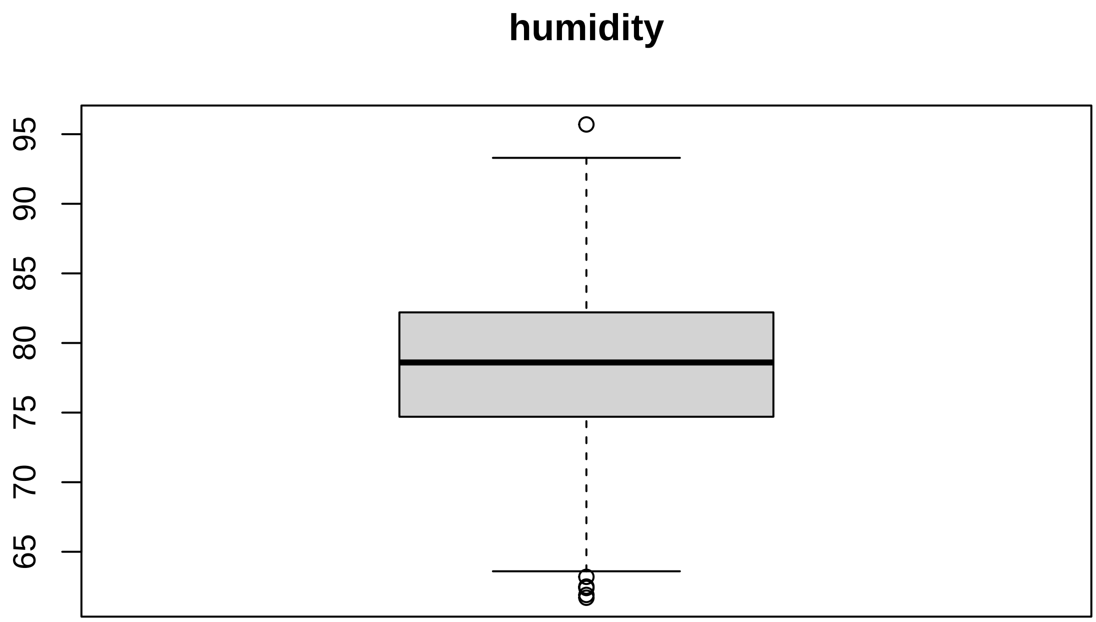

```
boxplot(Rainfall$feelslike)
title("feelslike")
```



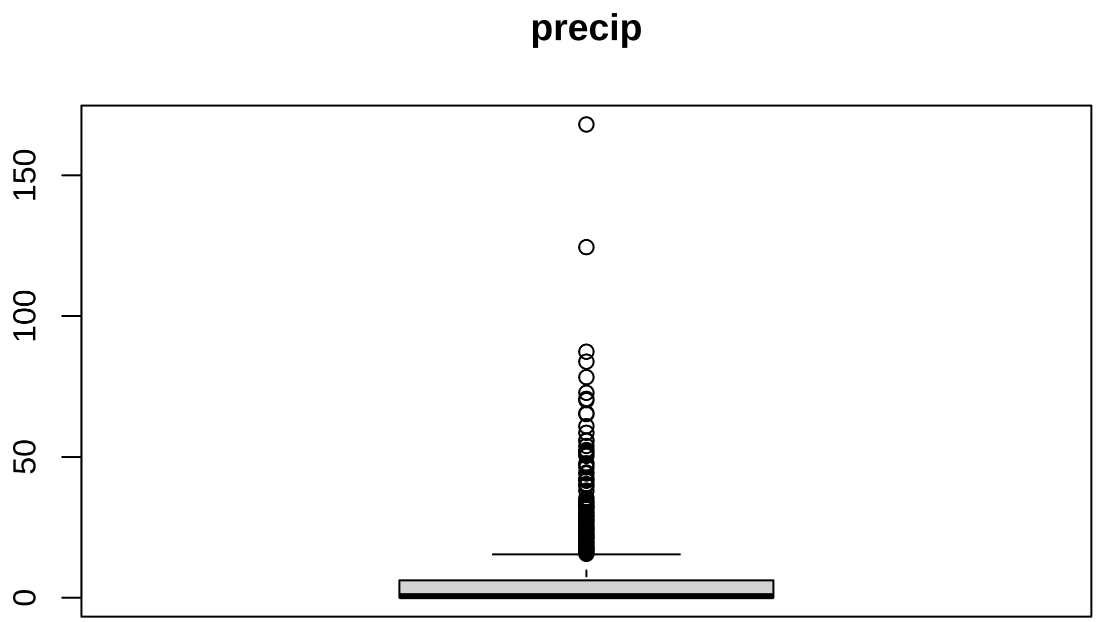

```r
boxplot(Rainfall$dew)
title("dew")
```



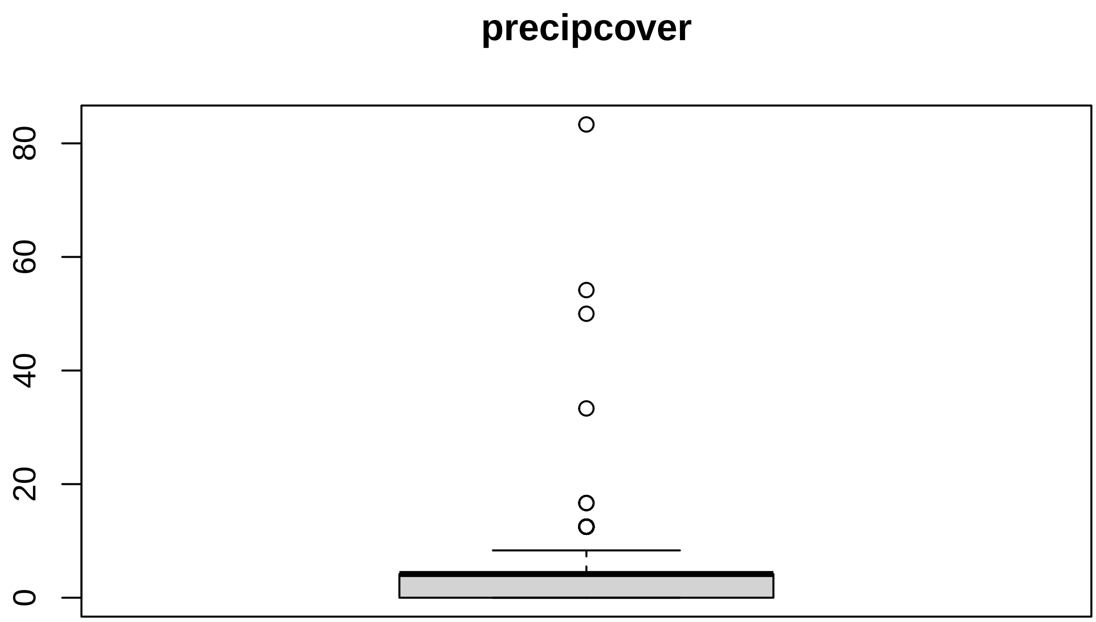

```r
boxplot(Rainfall$humidity)
title("humidity")
```



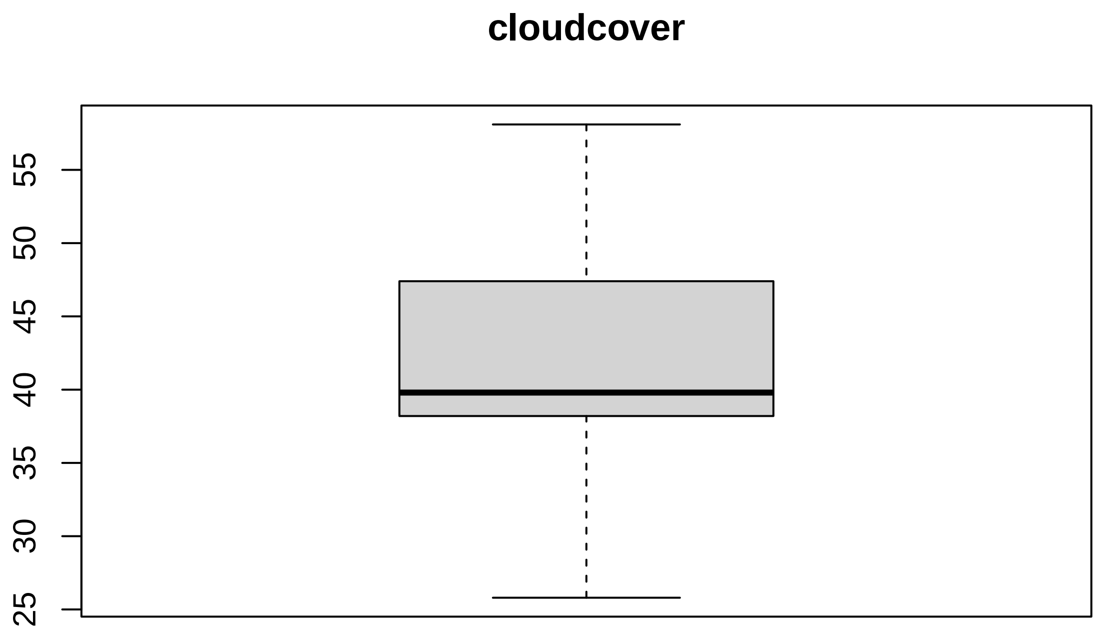

```r
boxplot(Rainfall$precip)
title("precip")
```



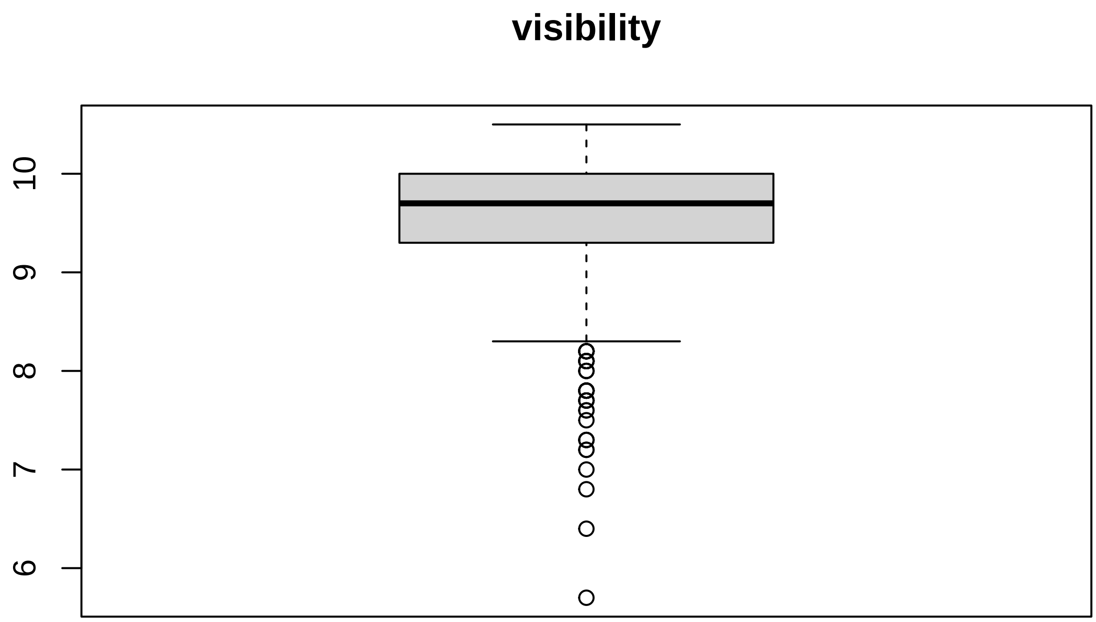

```
boxplot(Rainfall$precipcover)
title("precipcover")
```



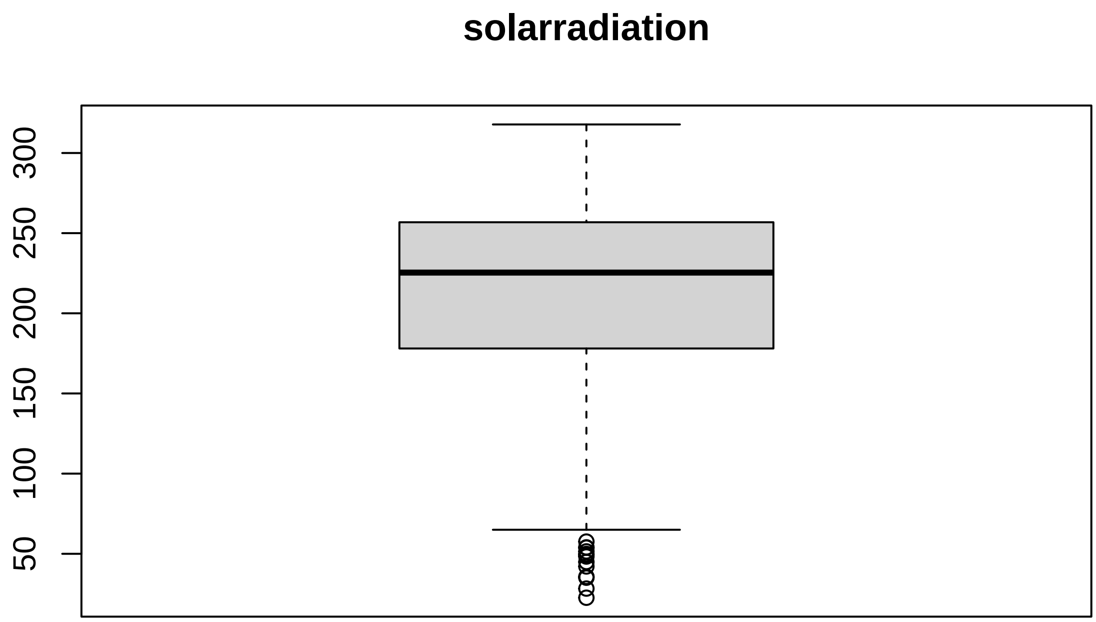

```r
boxplot(Rainfall$cloudcover)
title("cloudcover")
```



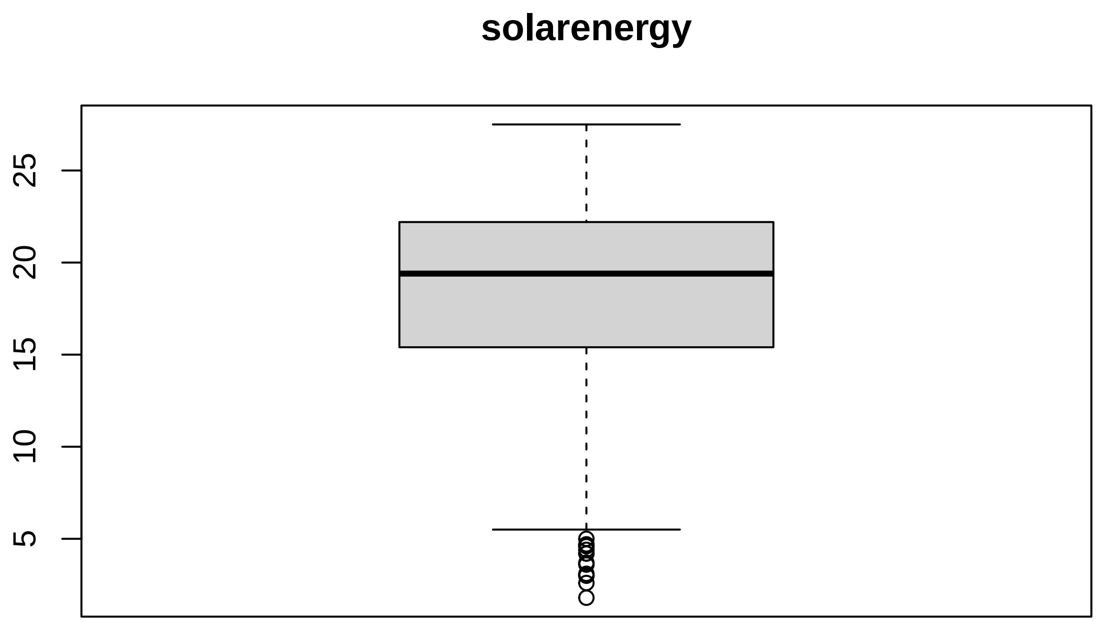

```r
boxplot(Rainfall$visibility)
title("visibility")
```




```
boxplot(Rainfall$solarradiation)
title("solarradiation")
```




```r
boxplot(Rainfall$solarenergy)
title("solarenergy")
```



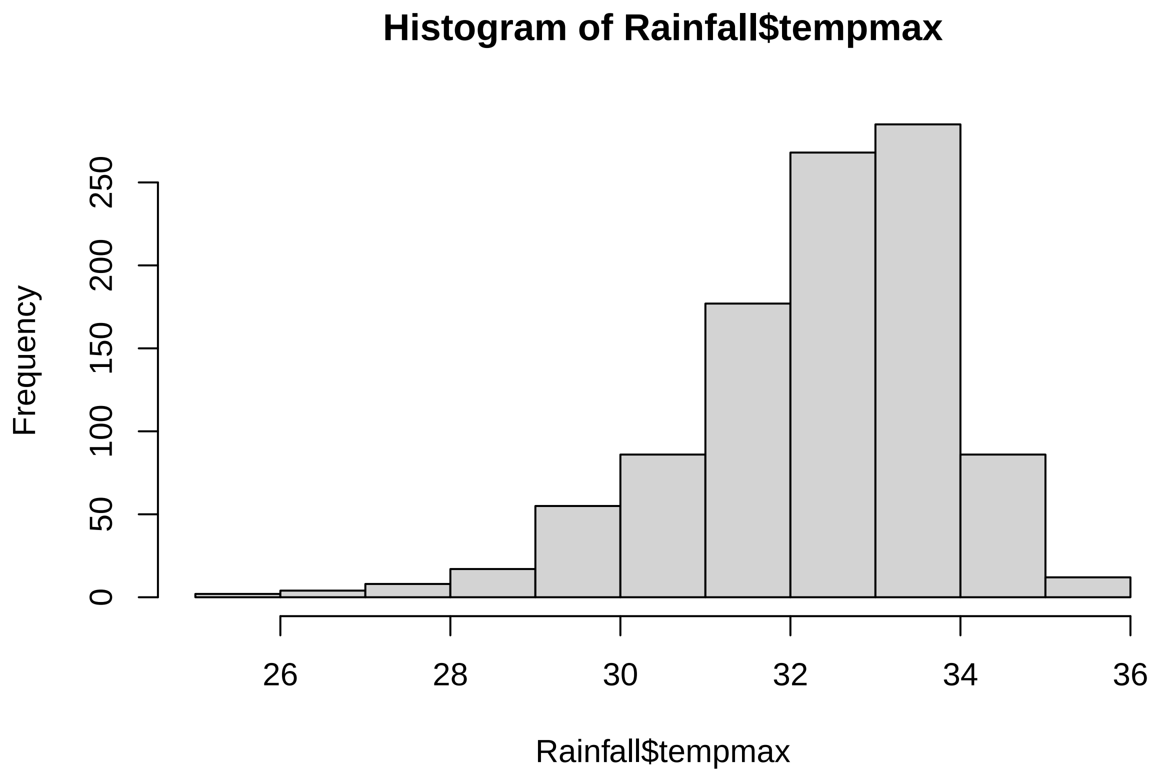
##Check duplication

```r
sum(duplicated(Rainfall))
```

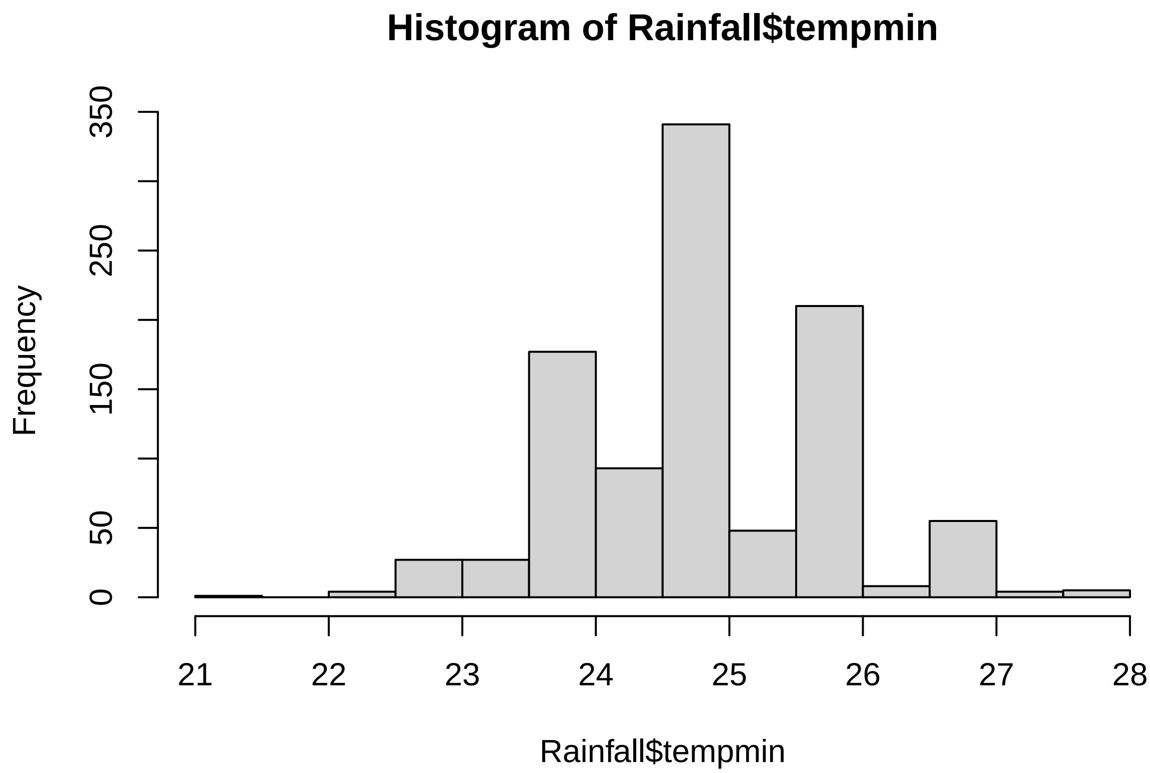
```
## [1] 0
```

#skewness

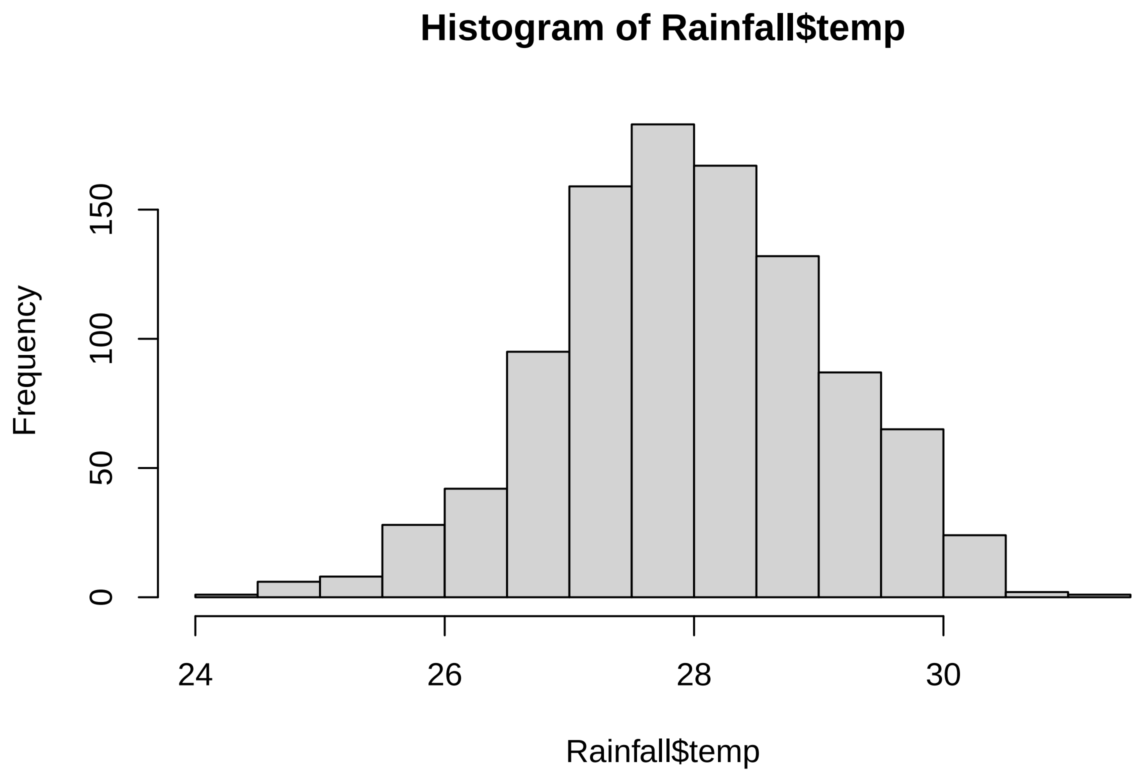
```r
hist(Rainfall$tempmax)
```



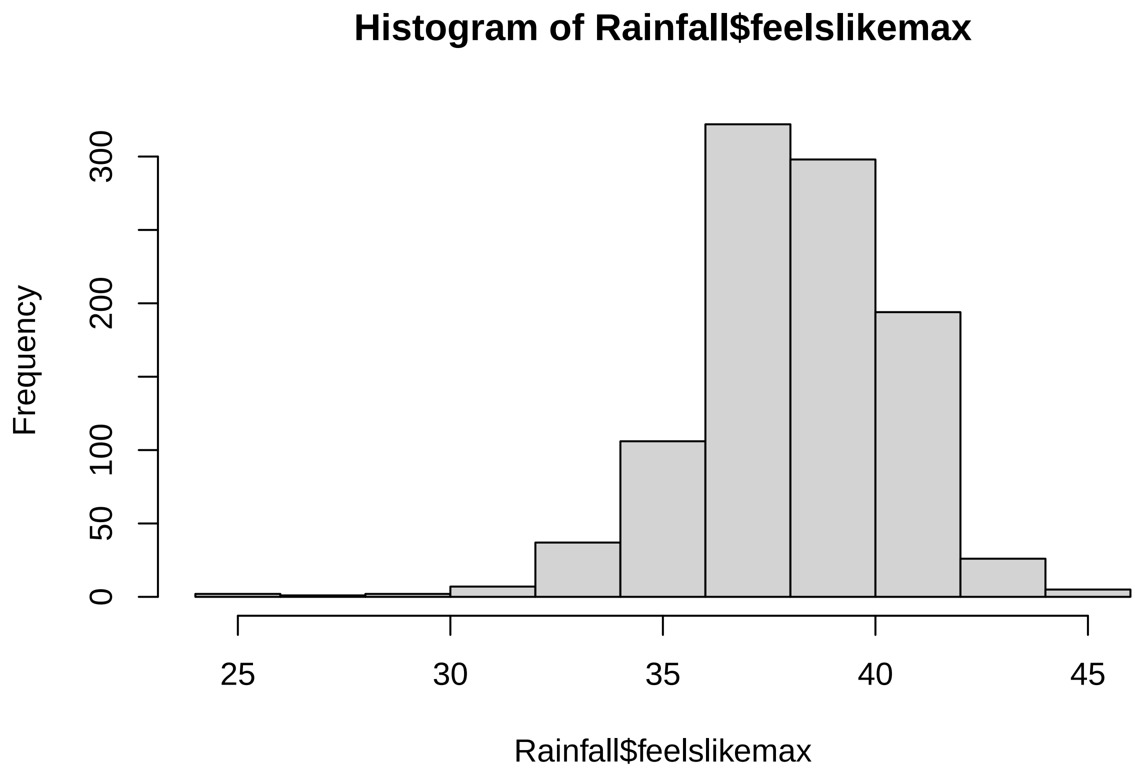

```
hist(Rainfall$tempmin)
```



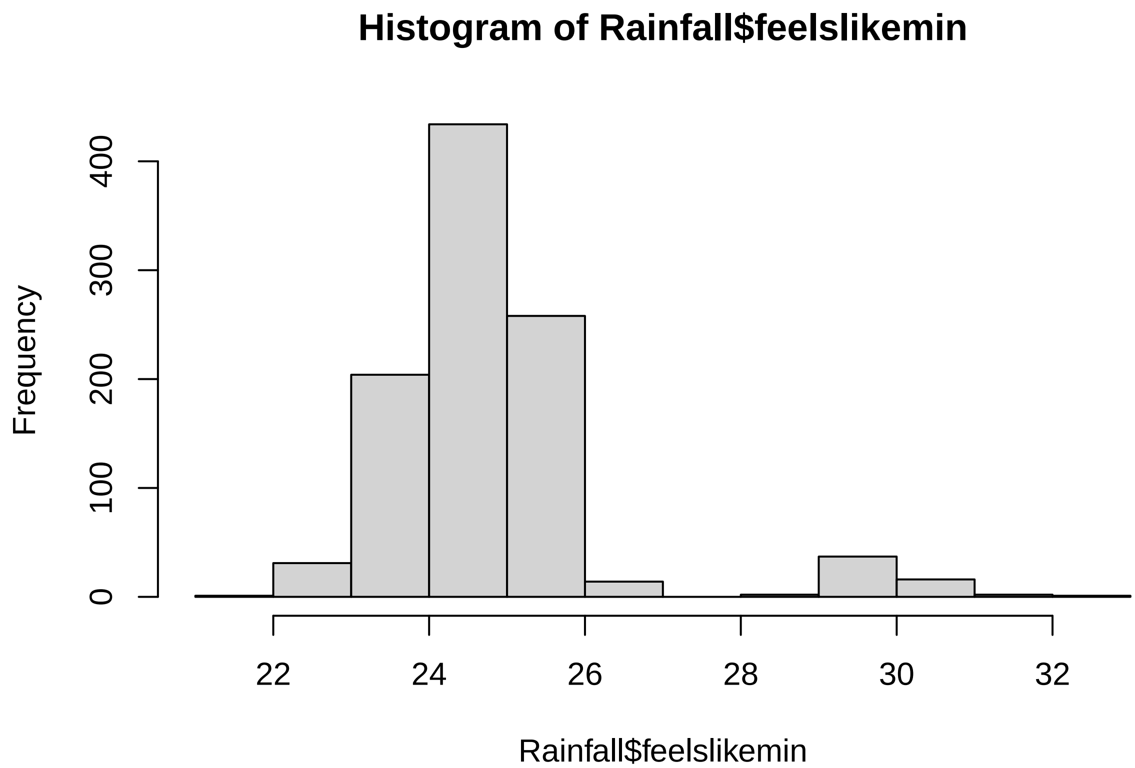

```r
hist(Rainfall$temp)
```



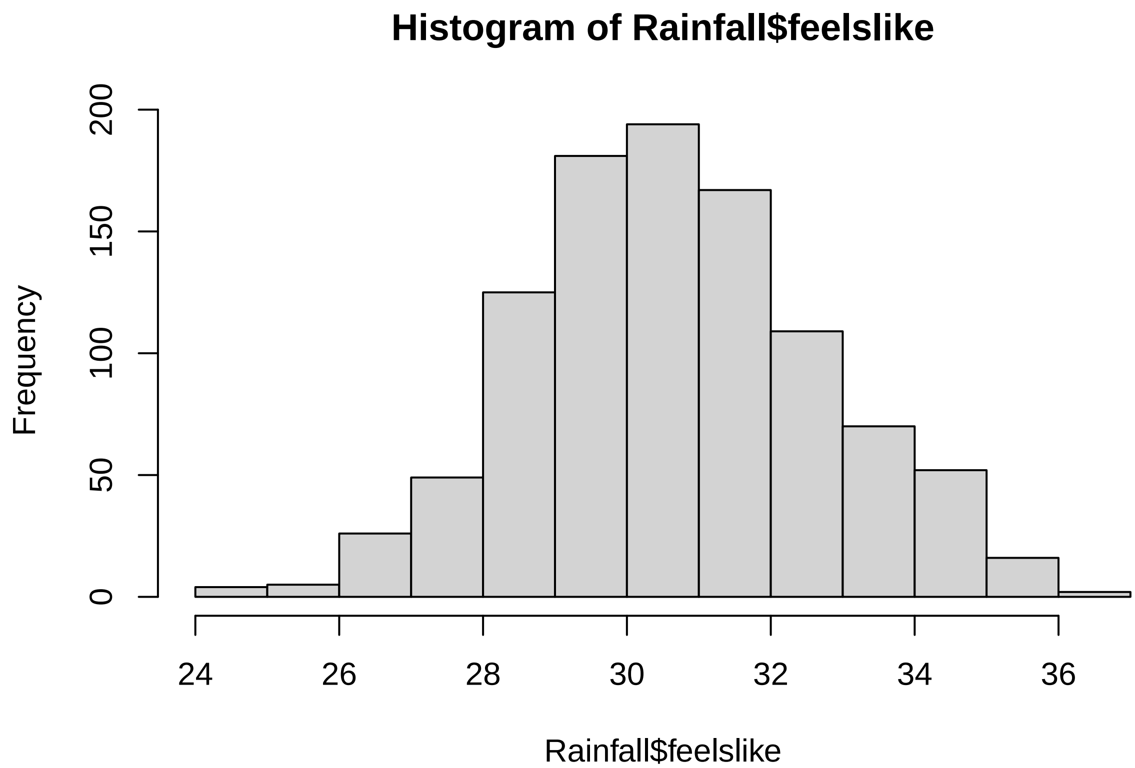

```
hist(Rainfall$feelslikemax)
```



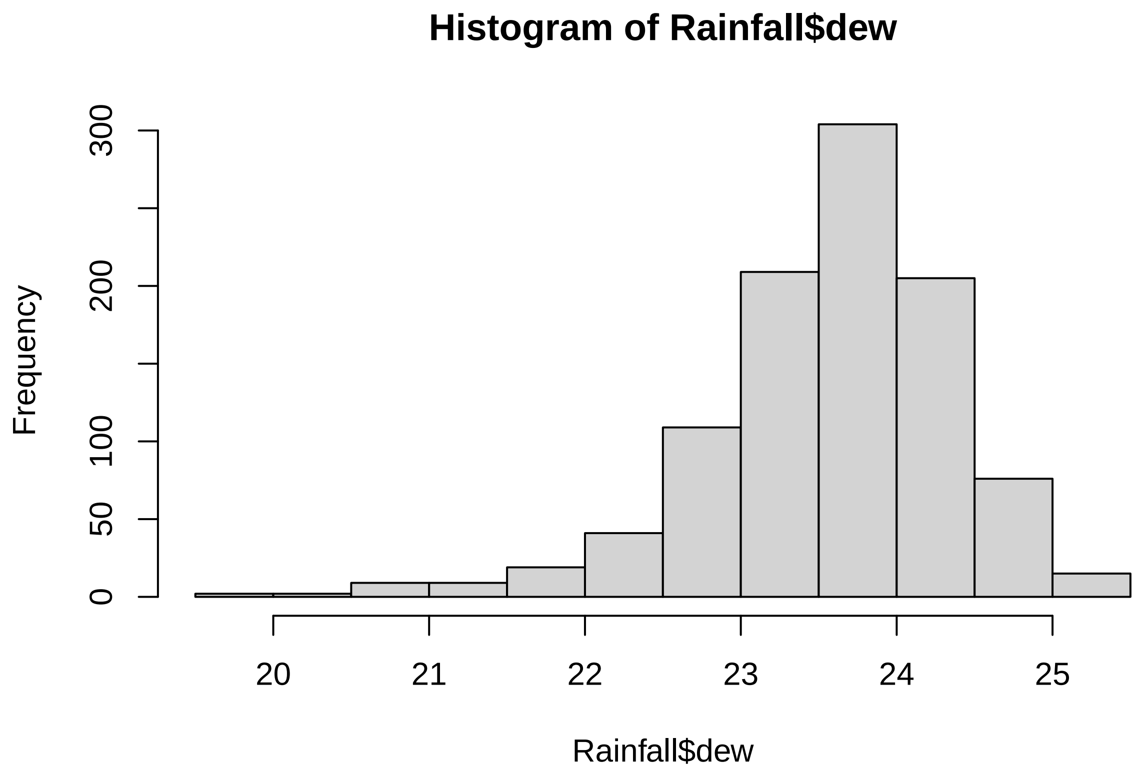

```r
hist(Rainfall$feelslikemin)
```



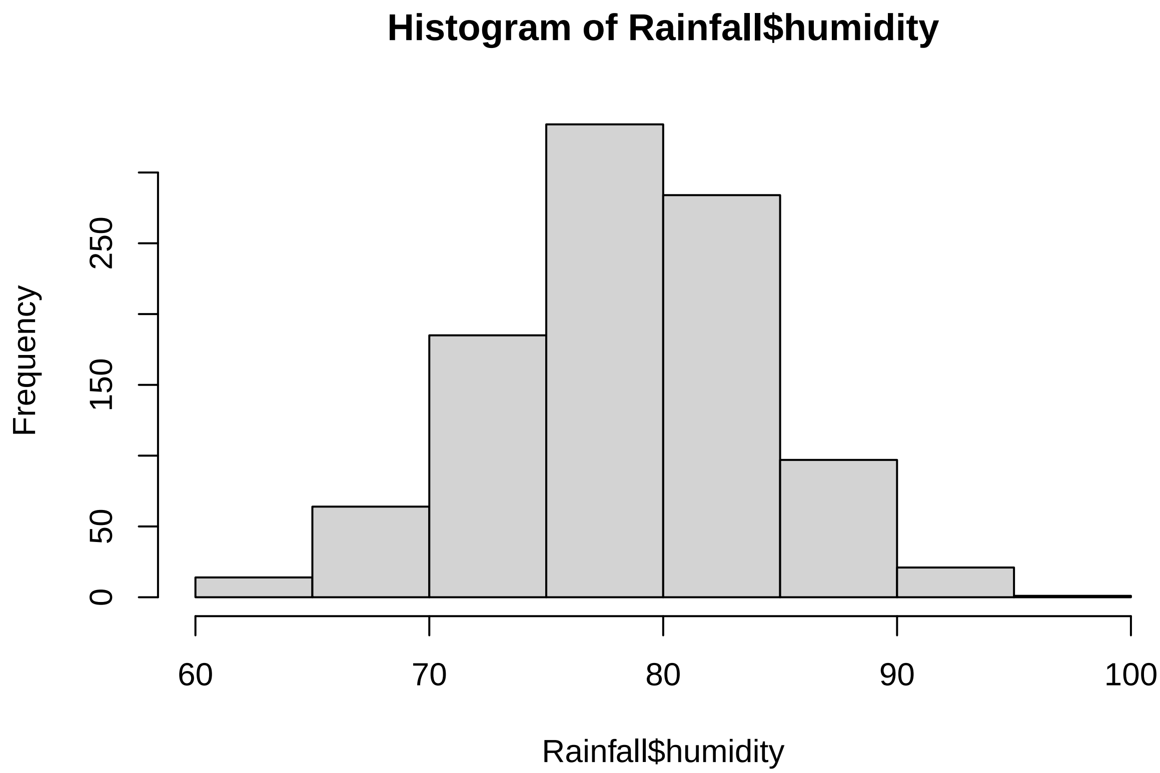

```r
hist(Rainfall$feelslike)
```



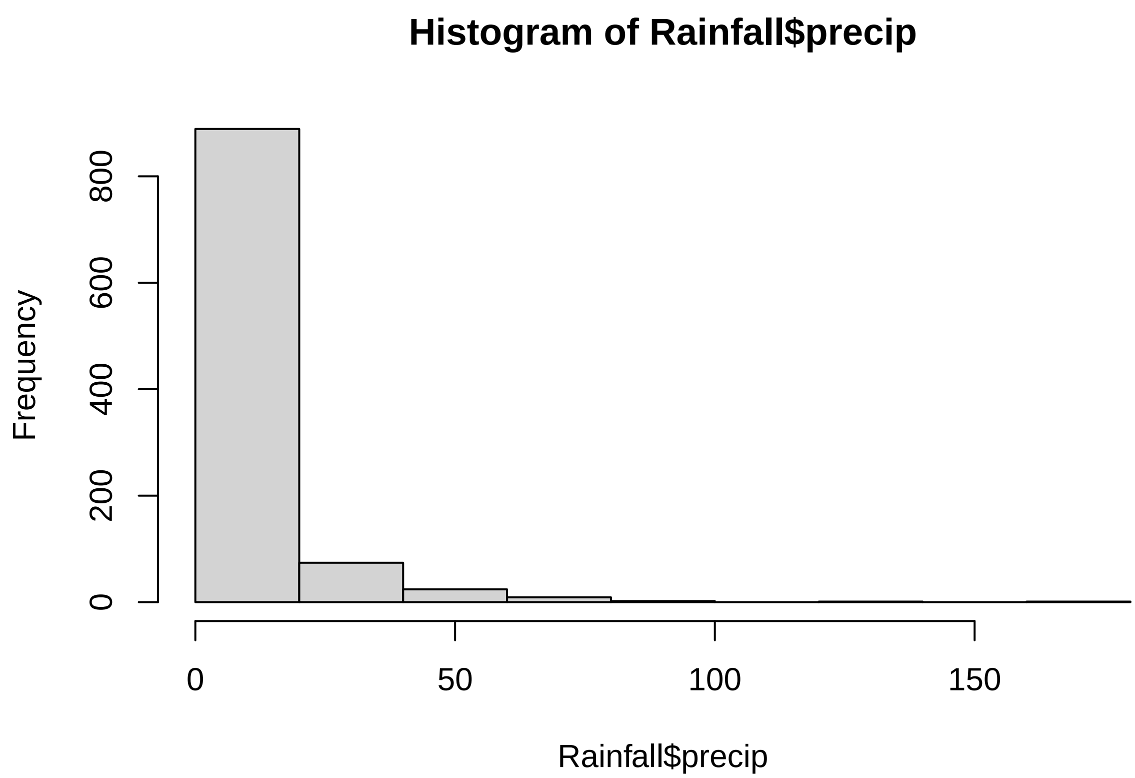

```r
hist(Rainfall$dew)
```



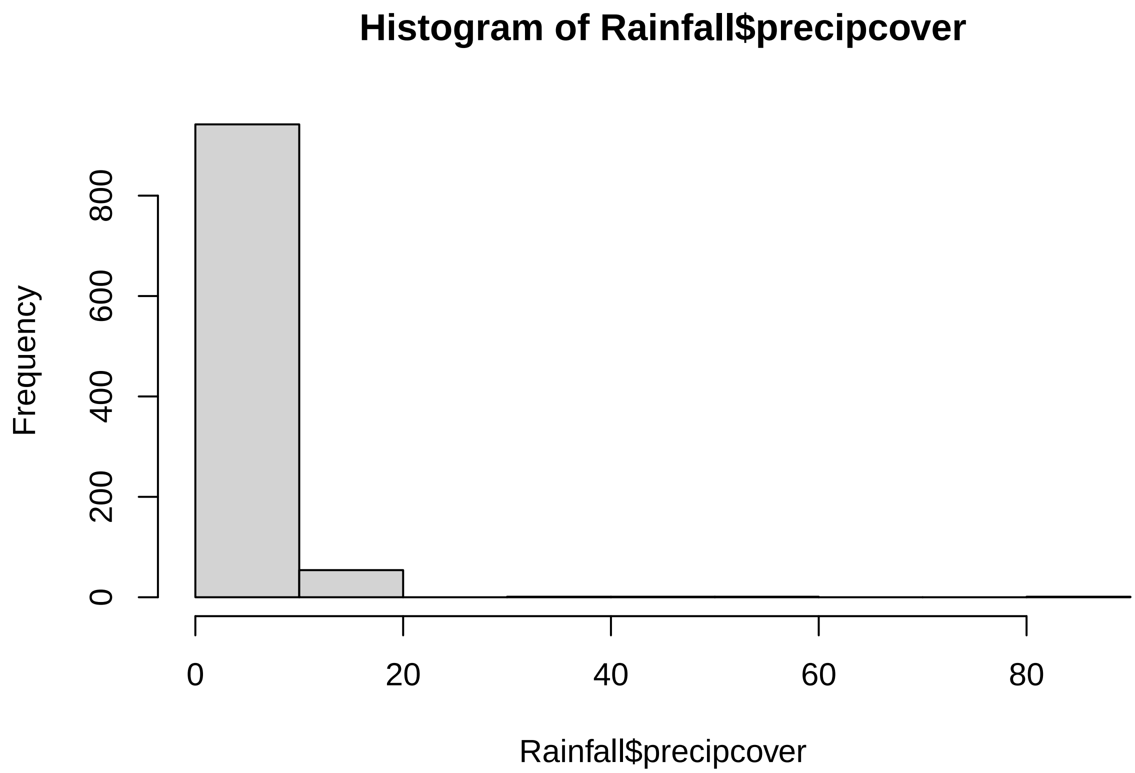

```
hist(Rainfall$humidity)
```



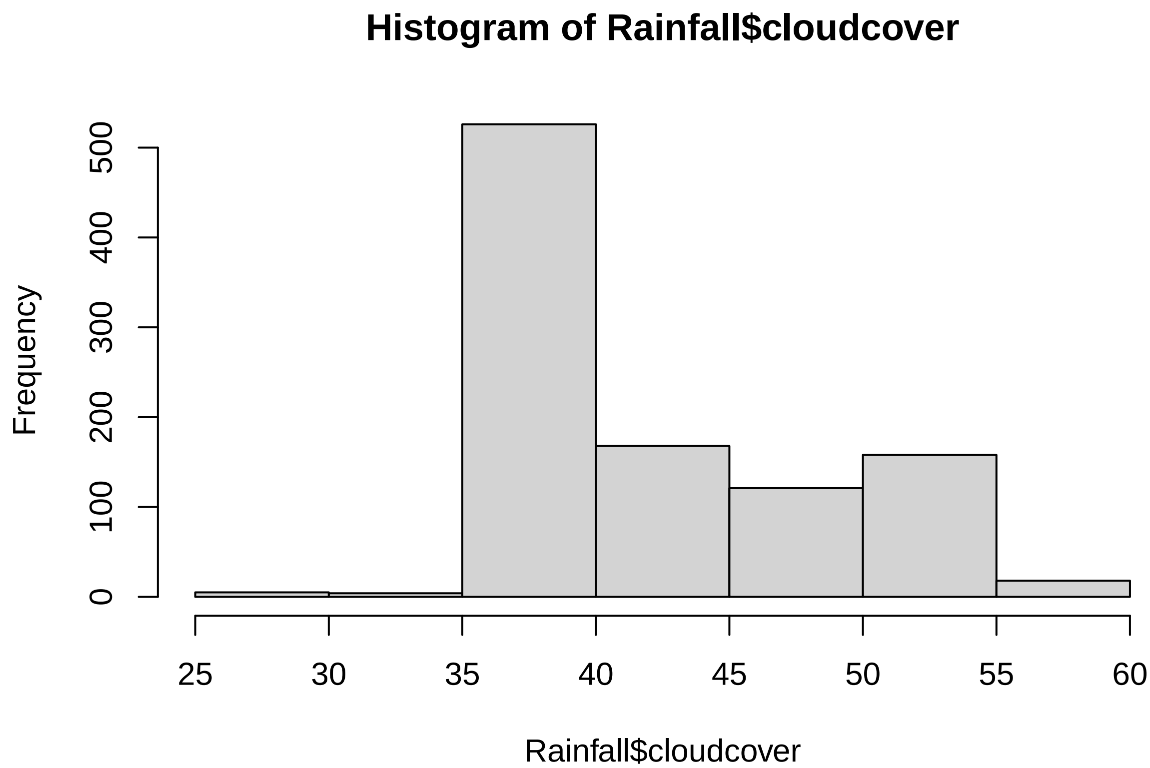

```
hist(Rainfall$precip)
```



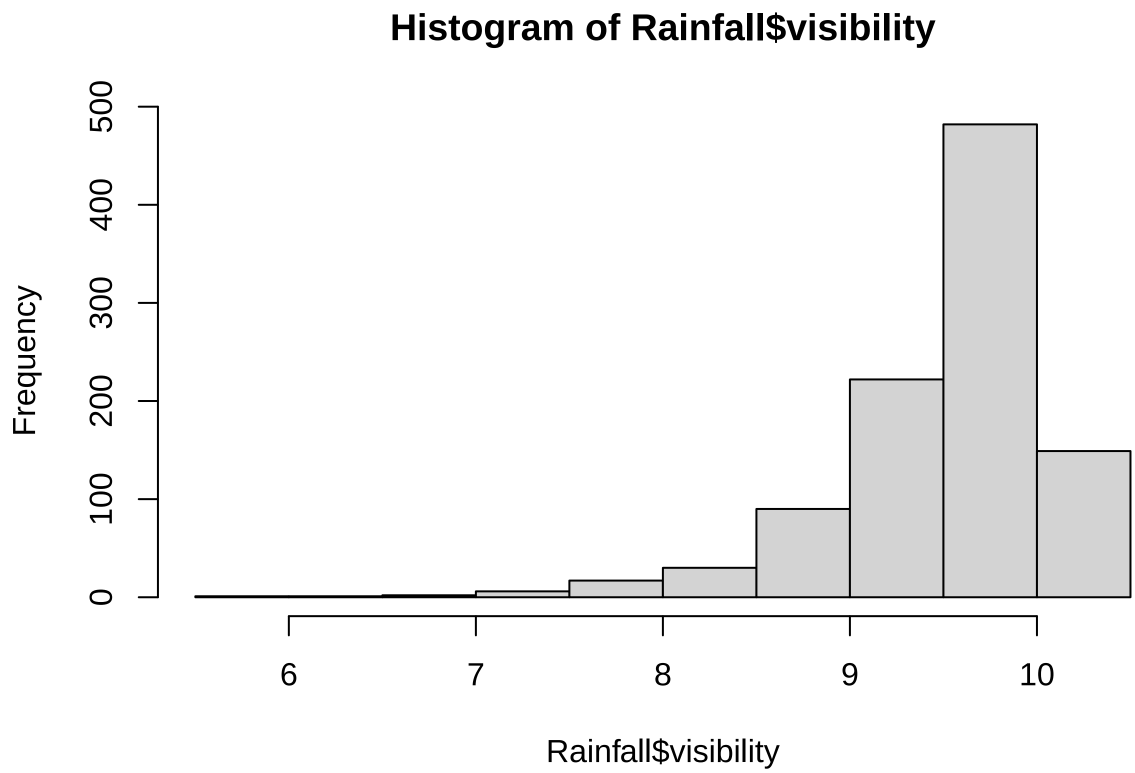

```r
hist(Rainfall$precipcover)
```



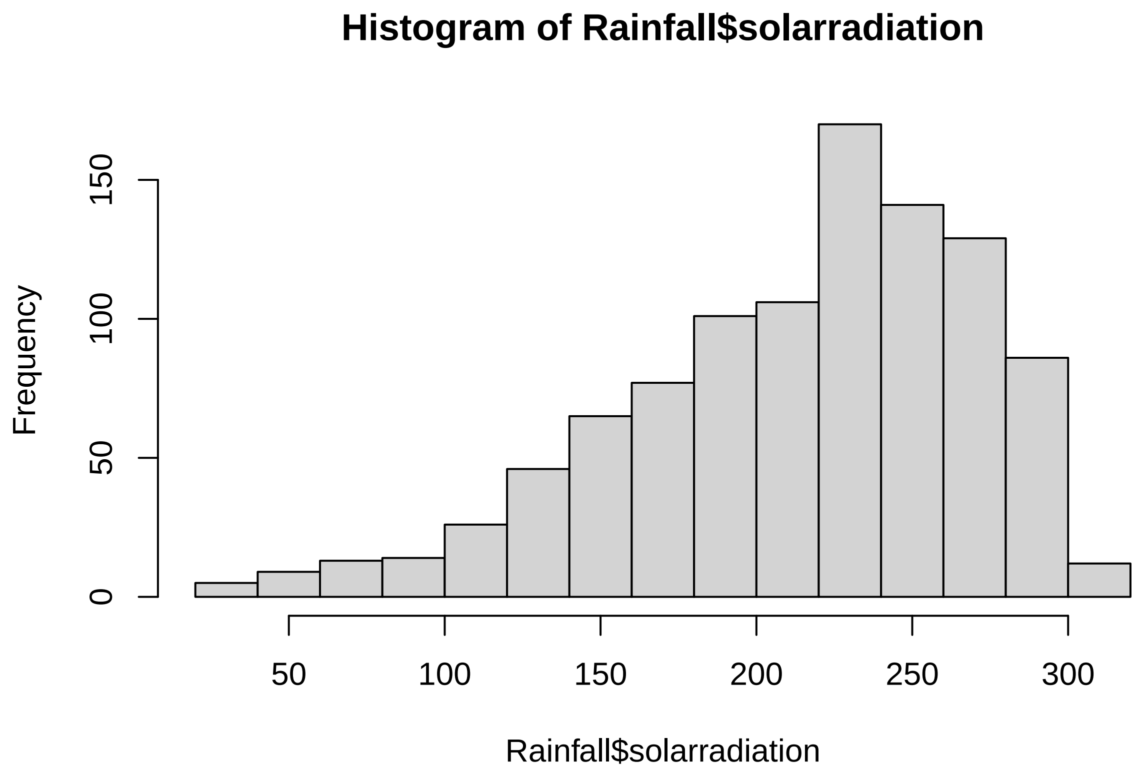

```r
hist(Rainfall$cloudcover)
```



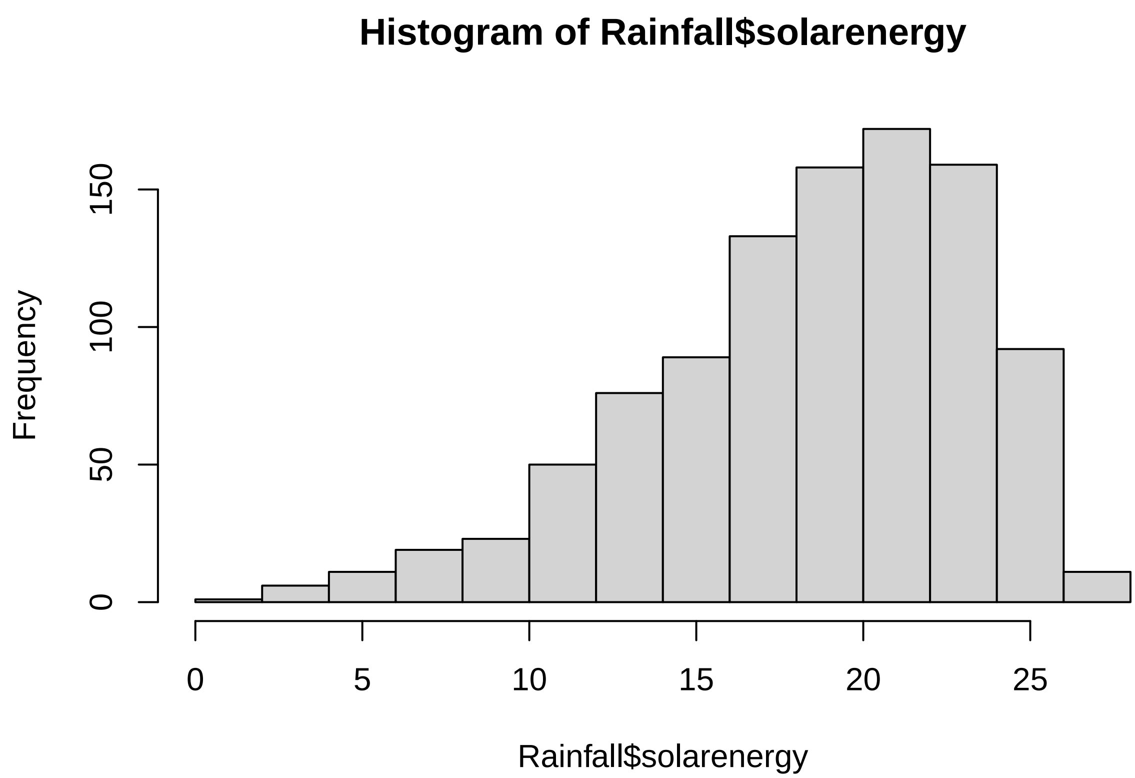

```
hist(Rainfall$visibility)
```




```
hist(Rainfall$solarradiation)
```




```
hist(Rainfall$solarenergy)
```



#skewness test

```r
install.packages("moments")
```

```
## Installing package into 'C:/Users/XIN WEI/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'moments' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\XIN WEI\AppData\Local\Temp\RtmpSK8f2W\downloaded_packages
```

```r
library(moments)
skewness(Rainfall$tempmax,na.rm=TRUE)
```

```
## [1] -0.9449293
```

```r
skewness(Rainfall$tempmin,na.rm=TRUE)
```

```
## [1] 0.1149611
```

```r
skewness(Rainfall$temp,na.rm=TRUE)
```

```
## [1] -0.1375686
```

```r
skewness(Rainfall$feelslikemax,na.rm=TRUE)
```

```
## [1] -0.6173491
```

```r
skewness(Rainfall$feelslikemin,na.rm=TRUE)
```

```
## [1] 1.993798
```

```r
skewness(Rainfall$feelslike,na.rm=TRUE)
```

```
## [1] 0.1003294
```

```r
skewness(Rainfall$dew,na.rm=TRUE)
```

```
## [1] -1.008102
```

```r
skewness(Rainfall$humidity,na.rm=TRUE)
```

```
## [1] -0.1757896
```

```r
skewness(Rainfall$precip,na.rm=TRUE)
```

```
## [1] 4.22407
```

```r
skewness(Rainfall$precipcover,na.rm=TRUE)
```

```
## [1] 6.209487
```

```r
skewness(Rainfall$cloudcover,na.rm=TRUE)
```

```
## [1] 0.8282789
```

```r
skewness(Rainfall$visibility,na.rm=TRUE)
```

```
## [1] -1.892846
```

```r
skewness(Rainfall$solarradiation,na.rm=TRUE)
```

```
## [1] -0.7626326
```

```r
skewness(Rainfall$solarenergy,na.rm=TRUE)
```

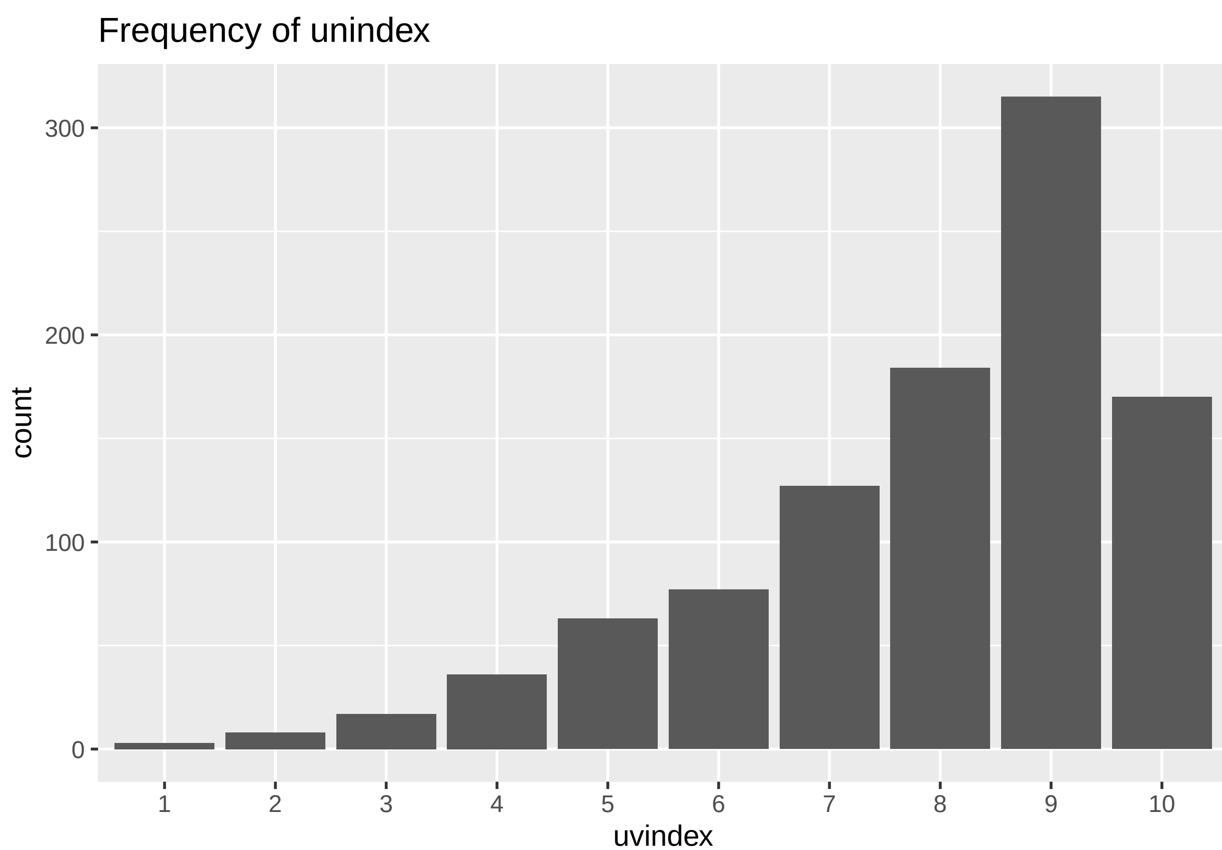
```
## [1] -0.7636563
```

#Categorical variable

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```r
uvi<-ggplot(Rainfall,aes(x=uvindex))+geom_bar()
uvi + ggtitle("Frequency of unindex")
```



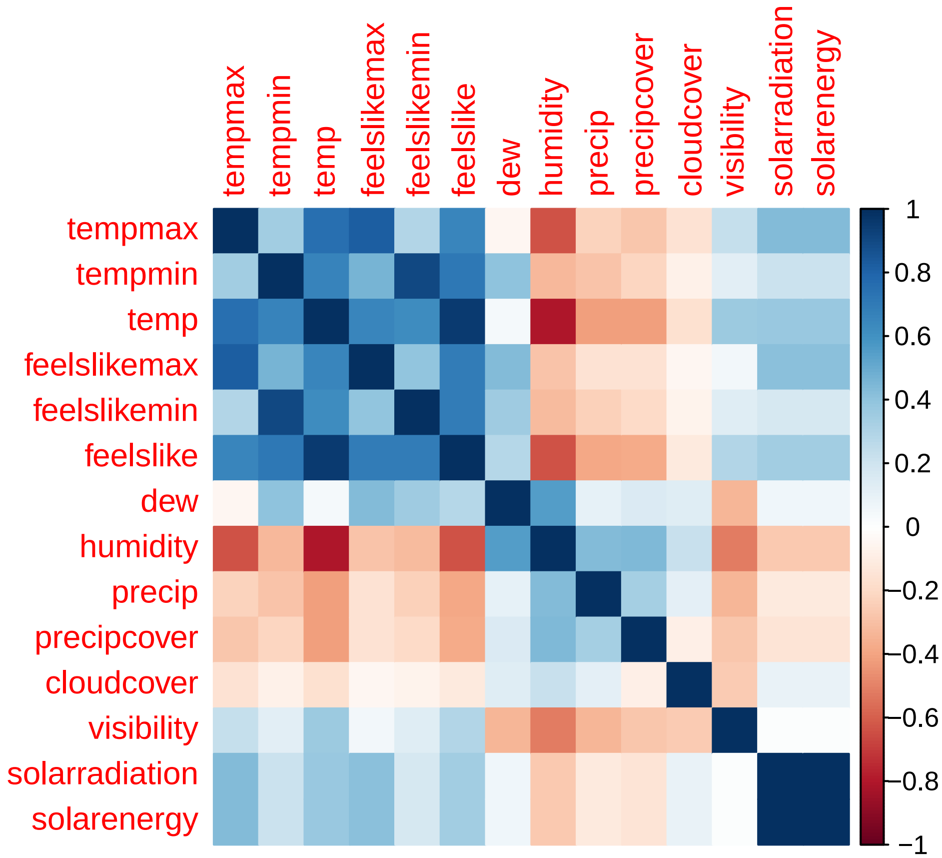
#correlation

```r
Rainfall_sub <- subset(Rainfall, select = -c(description, uvindex))
cr_Rainfall <- cor(Rainfall_sub)
cr_Rainfall
```

```
##                  tempmax     tempmin        temp feelslikemax feelslikemin
## tempmax       1.00000000  0.34628627  0.75697724   0.82183941    0.2988673
## tempmin       0.34628627  1.00000000  0.66498921   0.46399103    0.9090924
## temp          0.75697724  0.66498921  1.00000000   0.65925566    0.6214853
## feelslikemax  0.82183941  0.46399103  0.65925566   1.00000000    0.3984416
## feelslikemin  0.29886734  0.90909242  0.62148532   0.39844155    1.0000000
## feelslike     0.65558517  0.71652887  0.95234958   0.69867355    0.6901790
## dew          -0.04904635  0.40986190  0.04211583   0.43098257    0.3512328
## humidity     -0.63512575 -0.32344322 -0.80675532  -0.28192083   -0.3194219
## precip       -0.22823523 -0.28753507 -0.41880522  -0.15148459   -0.2378546
## precipcover  -0.27616976 -0.21653887 -0.41953149  -0.15795598   -0.1992029
```

```
## cloudcover     -0.15192631 -0.07385342 -0.16376114  -0.04422532   -0.0693215
## visibility      0.23267736  0.12565522  0.36044288   0.05390195    0.1395985
## solarradiation  0.43849395  0.21426365  0.37407025   0.41577111    0.1749074
## solarenergy     0.43904078  0.21577451  0.37602754   0.41571567    0.1762186
##                 feelslike         dew   humidity     precip precipcover
## tempmax         0.6555852 -0.04904635 -0.6351258 -0.2282352 -0.27616976
## tempmin         0.7165289  0.40986190 -0.3234432 -0.2875351 -0.21653887
## temp            0.9523496  0.04211583 -0.8067553 -0.4188052 -0.41953149
## feelslikemax    0.6986736  0.43098257 -0.2819208 -0.1514846 -0.15795598
## feelslikemin    0.6901790  0.35123283 -0.3194219 -0.2378546 -0.19920292
## feelslike       1.0000000  0.28192648 -0.6342491 -0.3804163 -0.37050155
## dew             0.2819265  1.00000000  0.5504398  0.1083176  0.15098855
## humidity       -0.6342491  0.55043979  1.0000000  0.4358605  0.44660758
## precip         -0.3804163  0.10831756  0.4358605  1.0000000  0.33492667
## precipcover    -0.3705015  0.15098855  0.4466076  0.3349267  1.00000000
## cloudcover     -0.1154345  0.13981211  0.2217531  0.1128520 -0.08821699
## visibility      0.2932208 -0.33233999 -0.5188545 -0.3352394 -0.27099911
## solarradiation  0.3422452  0.06210029 -0.2635237 -0.1103757 -0.14227183
## solarenergy     0.3440782  0.06136088 -0.2657426 -0.1127358 -0.14449900
##                 cloudcover  visibility solarradiation solarenergy
## tempmax        -0.15192631  0.23267736     0.43849395  0.43904078
## tempmin        -0.07385342  0.12565522     0.21426365  0.21577451
## temp           -0.16376114  0.36044288     0.37407025  0.37602754
## feelslikemax   -0.04422532  0.05390195     0.41577111  0.41571567
## feelslikemin   -0.06932150  0.13959848     0.17490744  0.17621857
## feelslike      -0.11543448  0.29322085     0.34224519  0.34407816
## dew             0.13981211 -0.33233999     0.06210029  0.06136088
## humidity        0.22175313 -0.51885448    -0.26352374 -0.26574257
## precip          0.11285197 -0.33523942    -0.11037574 -0.11273581
## precipcover    -0.08821699 -0.27099911    -0.14227183 -0.14449900
## cloudcover      1.00000000 -0.25781629     0.09395415  0.09431281
## visibility     -0.25781629  1.00000000     0.01863402  0.01947126
## solarradiation  0.09395415  0.01863402     1.00000000  0.99969905
## solarenergy     0.09431281  0.01947126     0.99969905  1.00000000
```

```r
corrplot(cr_Rainfall,method="color")
```



0.5 Bivariate Analysis

```r
library(tidycomm)
```

```
## Warning: package 'tidycomm' was built under R version 4.3.2
```

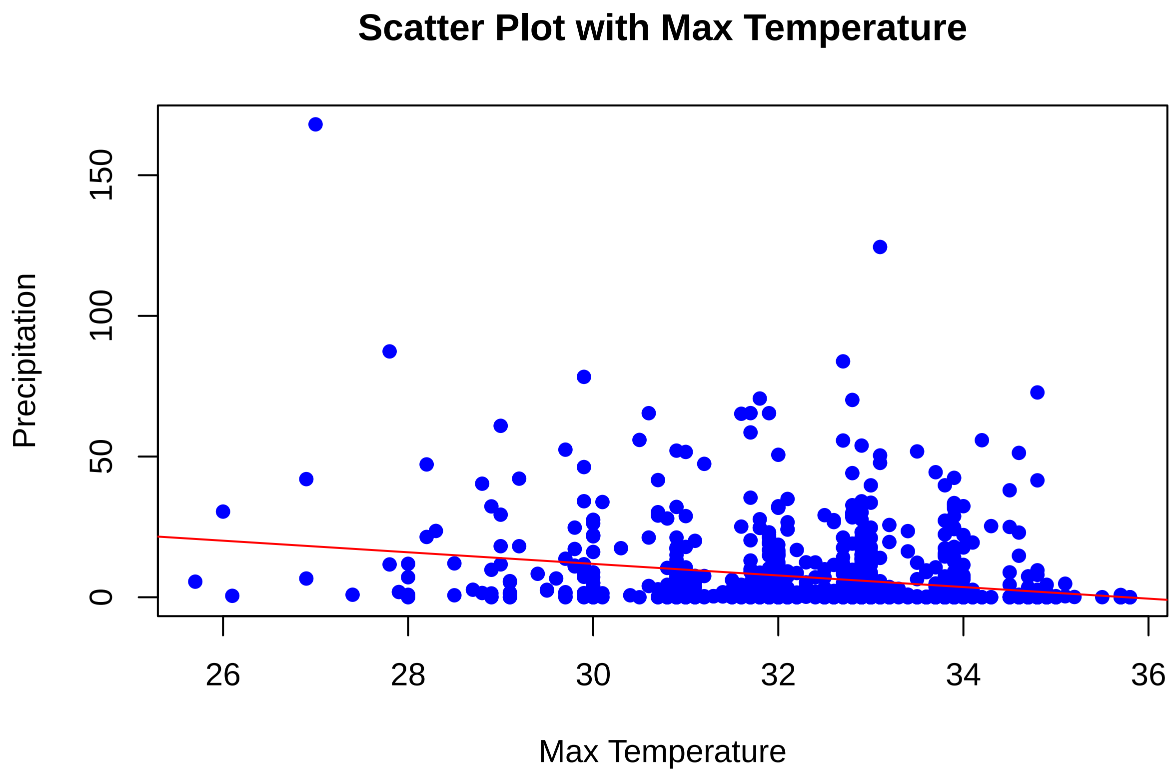
```r
correlate(Rainfall)
```

```
## # A tibble: 91 x 5
##    x       y                 r    df         p
##    <chr>   <chr>         <dbl> <int>     <dbl>
##  1 tempmax tempmin      0.346    998 1.51e- 29
##  2 tempmax temp         0.757    998 1.26e-186
##  3 tempmax feelslikemax 0.822    998 4.33e-246
##  4 tempmax feelslikemin 0.299    998 4.42e- 22
##  5 tempmax feelslike    0.656    998 7.01e-124
##  6 tempmax dew         -0.0490   998 1.21e-  1
##  7 tempmax humidity    -0.635    998 4.68e-114
##  8 tempmax precip      -0.228    998 2.77e- 13
##  9 tempmax precipcover -0.276    998 5.79e- 19
## 10 tempmax cloudcover  -0.152    998 1.39e-  6
## # i 81 more rows
```

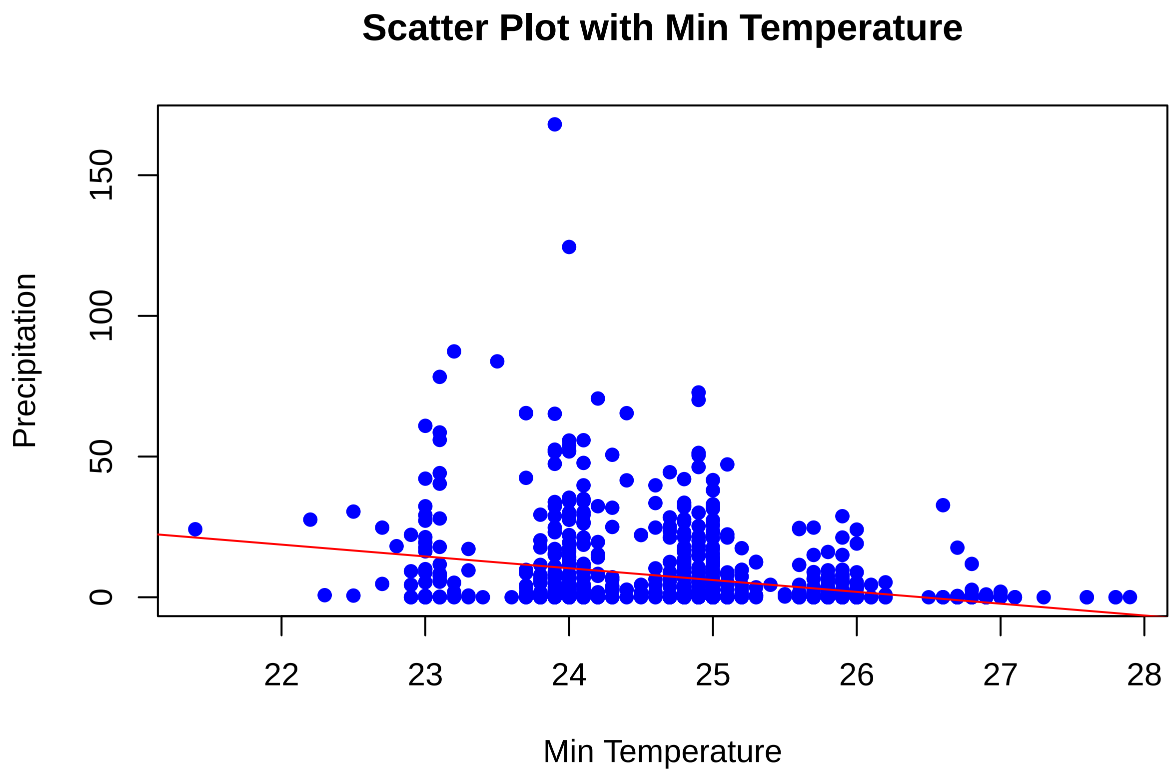
```r
correlate(Rainfall)%>%
  to_correlation_matrix()
```

```
## # A tibble: 14 x 15
##    r         tempmax tempmin    temp feelslikemax feelslikemin feelslike     dew
##    <chr>       <dbl>   <dbl>   <dbl>        <dbl>        <dbl>     <dbl>   <dbl>
##  1 tempmax     1       0.346   0.757        0.822        0.299     0.656 -0.0490
##  2 tempmin     0.346   1       0.665        0.464        0.909     0.717  0.410
##  3 temp        0.757   0.665   1            0.659        0.621     0.952  0.0421
##  4 feelslik~   0.822   0.464   0.659        1            0.398     0.699  0.431
##  5 feelslik~   0.299   0.909   0.621        0.398        1         0.690  0.351
##  6 feelslike   0.656   0.717   0.952        0.699        0.690     1      0.282
##  7 dew        -0.0490  0.410   0.0421       0.431        0.351     0.282  1
##  8 humidity   -0.635  -0.323  -0.807       -0.282       -0.319    -0.634  0.550
##  9 precip     -0.228  -0.288  -0.419       -0.151       -0.238    -0.380  0.108
## 10 precipco~ -0.276  -0.217  -0.420       -0.158       -0.199    -0.371  0.151
## 11 cloudcov~ -0.152  -0.0739 -0.164       -0.0442      -0.0693   -0.115  0.140
## 12 visibili~  0.233   0.126   0.360        0.0539       0.140     0.293 -0.332
## 13 solarrad~  0.438   0.214   0.374        0.416        0.175     0.342  0.0621
## 14 solarene~  0.439   0.216   0.376        0.416        0.176     0.344  0.0614
## # i 7 more variables: humidity <dbl>, precip <dbl>, precipcover <dbl>,
## #   cloudcover <dbl>, visibility <dbl>, solarradiation <dbl>, solarenergy <dbl>
```

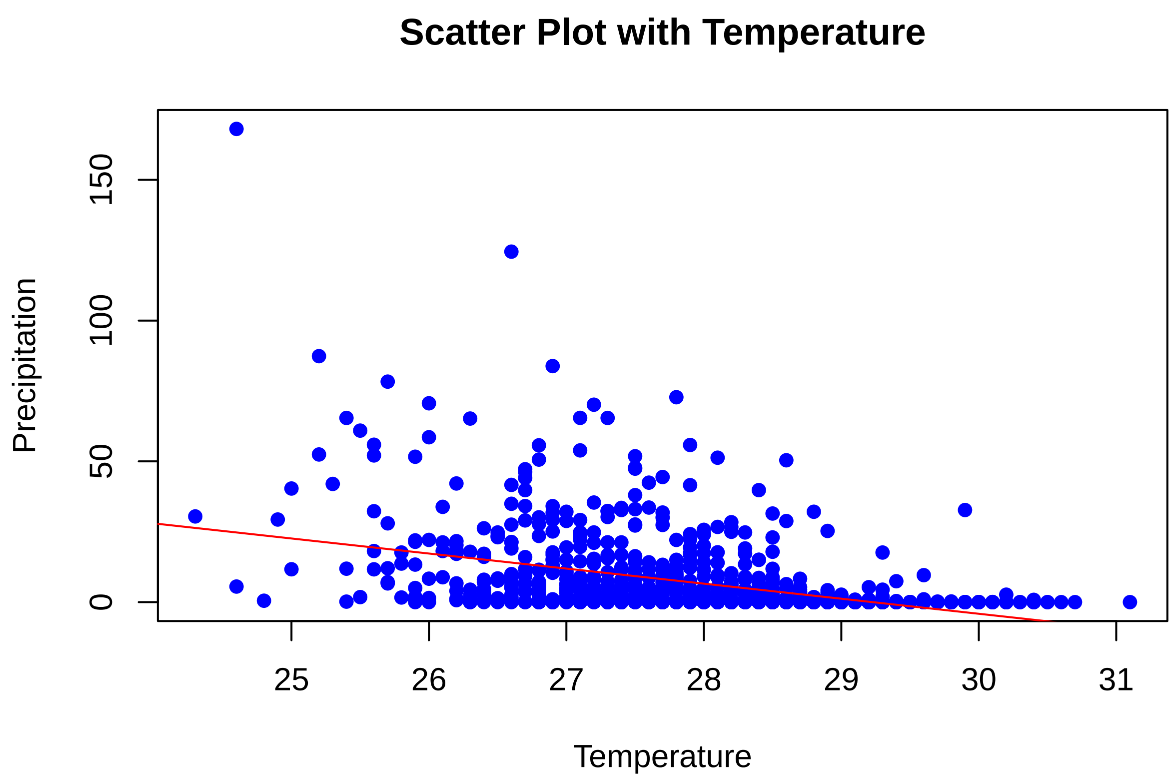
```r
plot(Rainfall$tempmax, Rainfall$precip,
     main="Scatter Plot with Max Temperature", xlab="Max Temperature",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$tempmax), col="red")
```



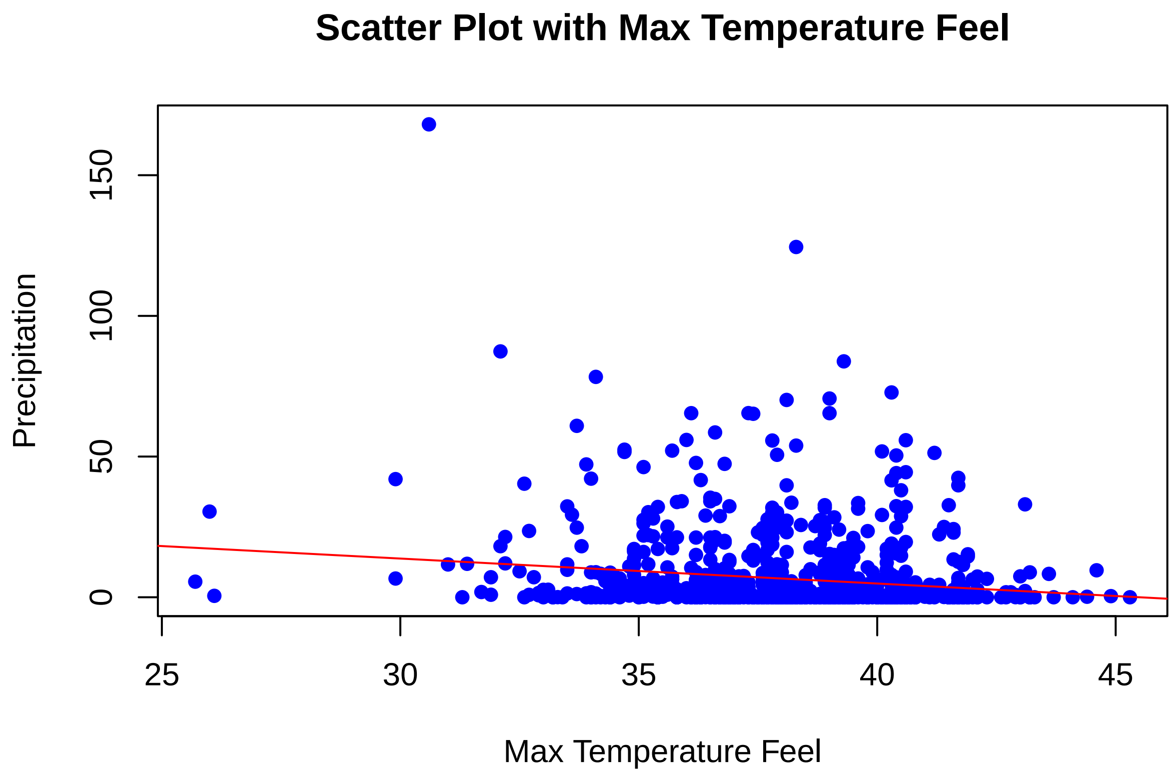

```r
plot(Rainfall$tempmin, Rainfall$precip,
     main="Scatter Plot with Min Temperature", xlab="Min Temperature",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$tempmin), col="red")
```



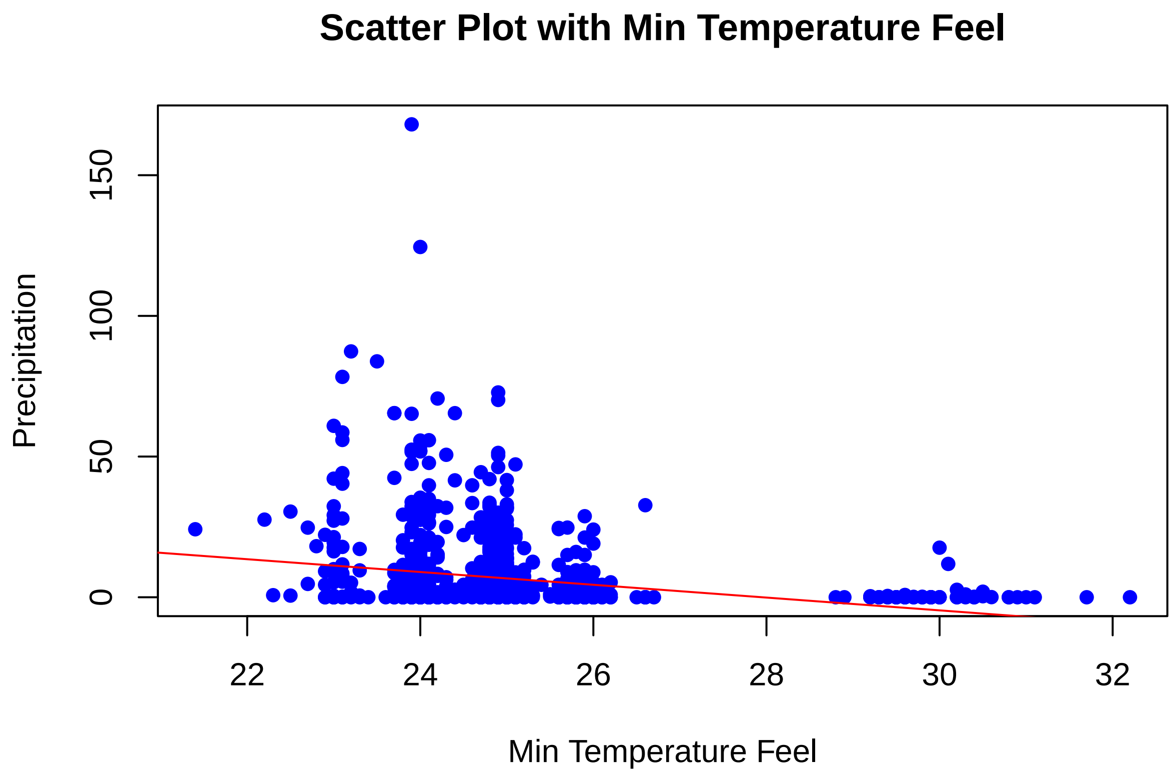

```r
plot(Rainfall$temp, Rainfall$precip,
     main="Scatter Plot with Temperature", xlab="Temperature",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$temp), col="red")
```



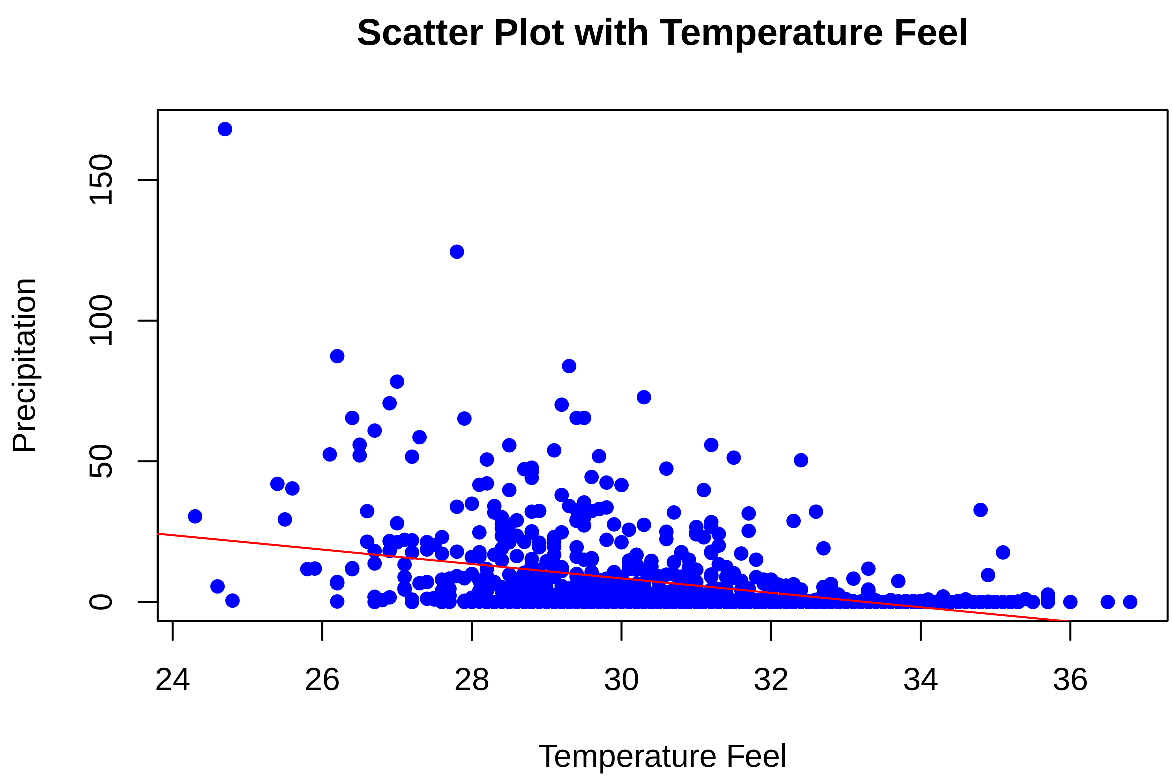

```r
plot(Rainfall$feelslikemax, Rainfall$precip,
     main="Scatter Plot with Max Temperature Feel", xlab="Max Temperature Feel",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$feelslikemax), col="red")
```



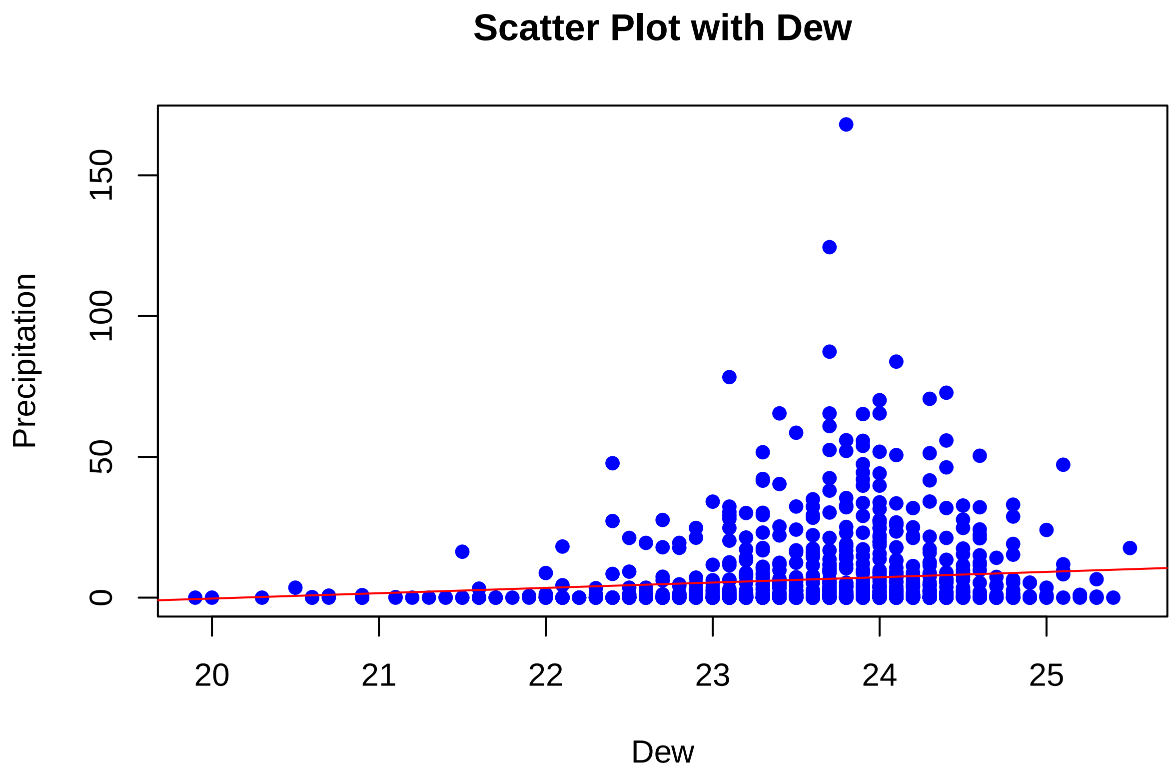

```r
plot(Rainfall$feelslikemin, Rainfall$precip,
     main="Scatter Plot with Min Temperature Feel", xlab="Min Temperature Feel",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$feelslikemin), col="red")
```



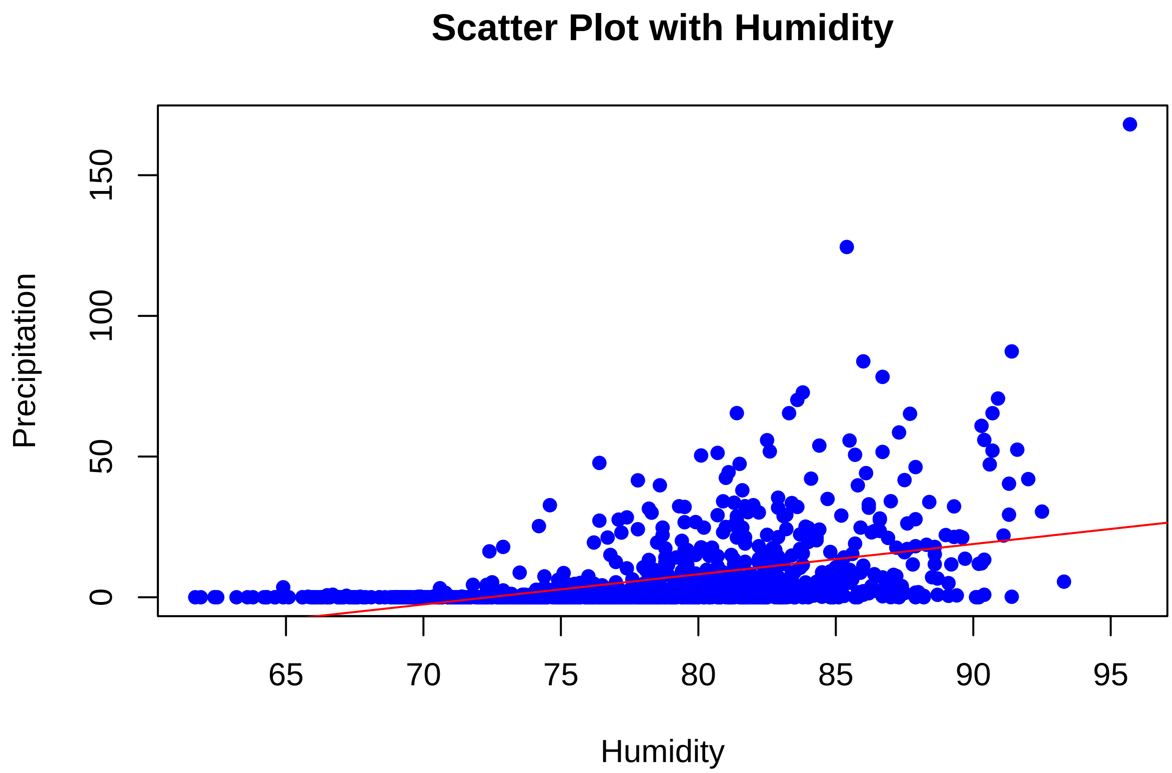

```r
plot(Rainfall$feelslike, Rainfall$precip,
     main="Scatter Plot with Temperature Feel", xlab="Temperature Feel",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$feelslike), col="red")
```



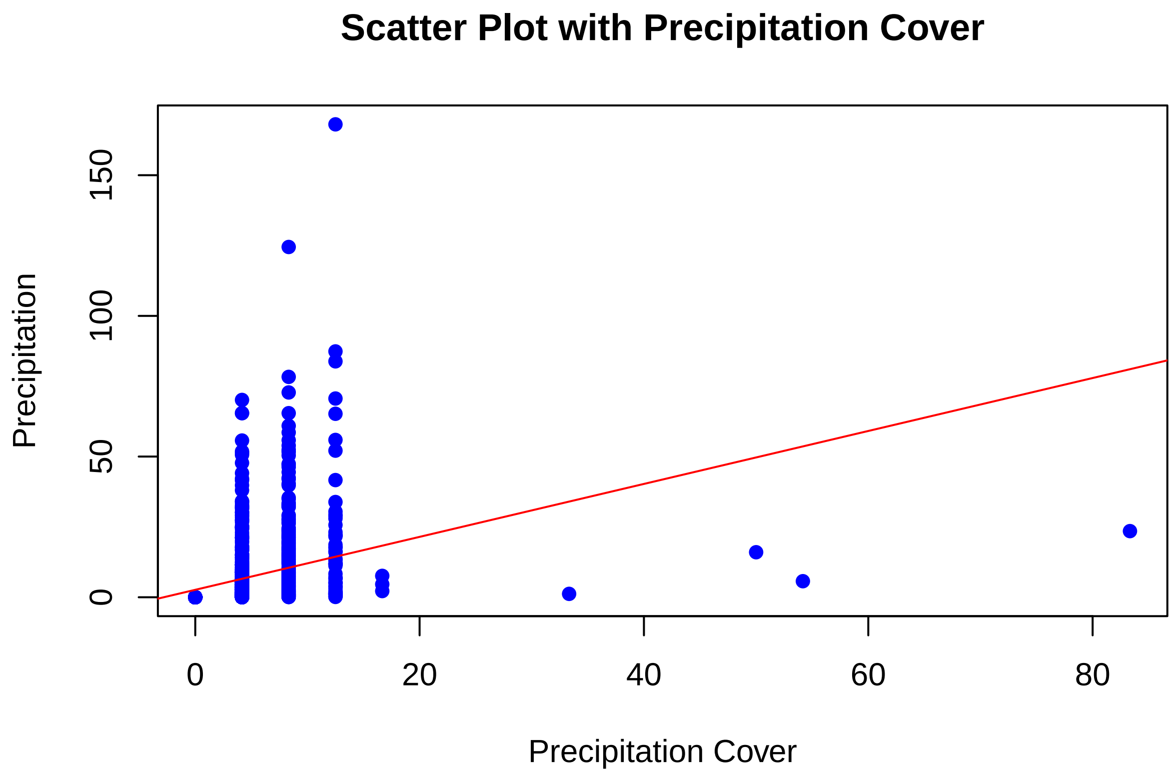

```r
plot(Rainfall$dew, Rainfall$precip,
     main="Scatter Plot with Dew", xlab="Dew",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$dew), col="red")
```



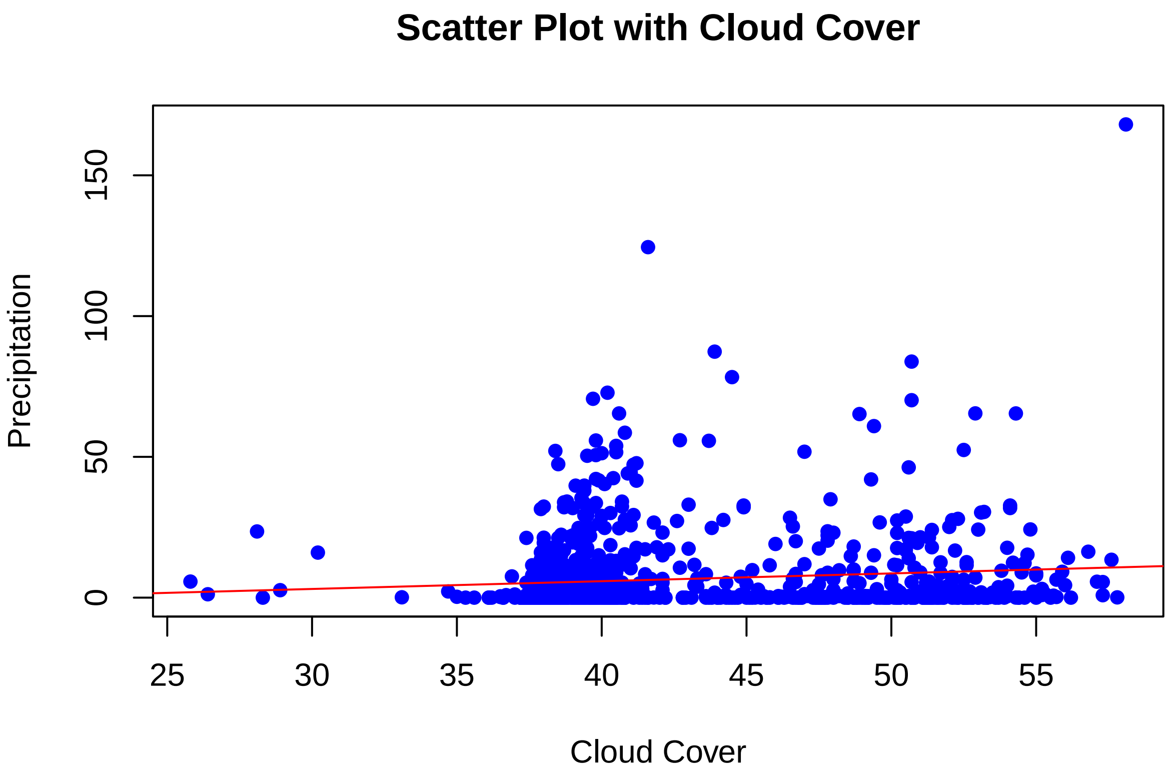

```r
plot(Rainfall$humidity, Rainfall$precip,
     main="Scatter Plot with Humidity", xlab="Humidity",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$humidity), col="red")
```



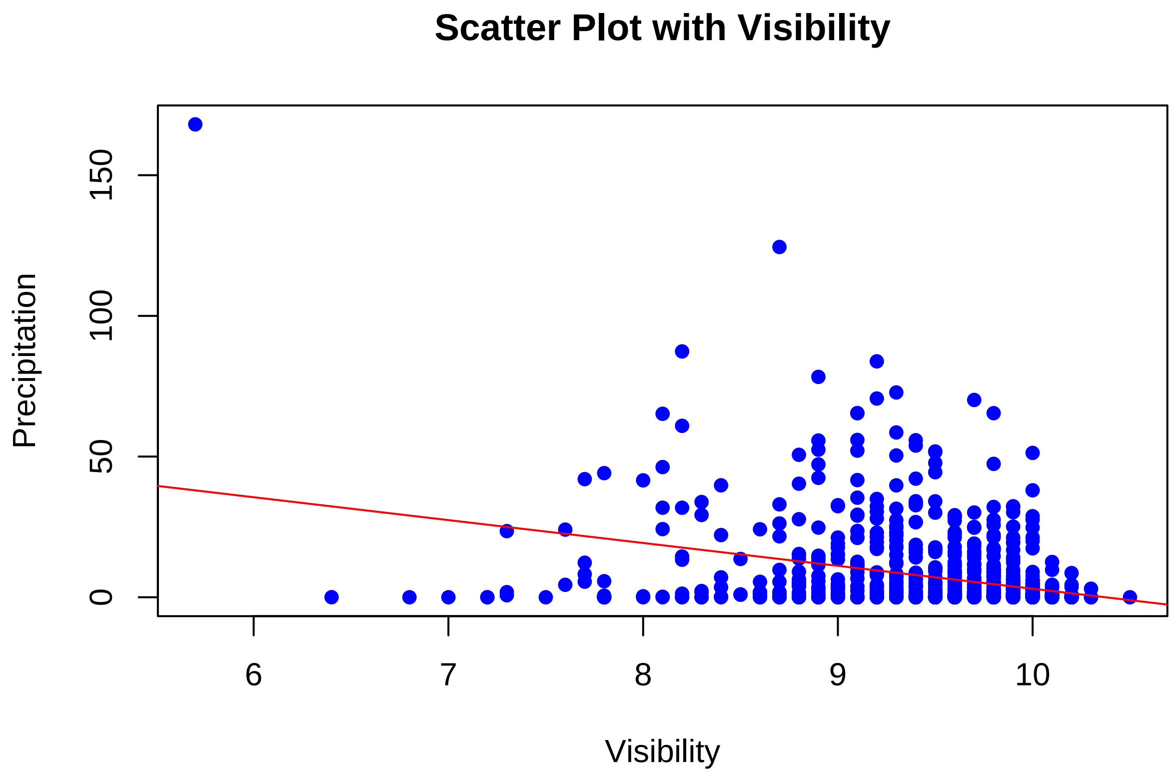

```
plot(Rainfall$precipcover, Rainfall$precip,
     main="Scatter Plot with Precipitation Cover", xlab="Precipitation Cover",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$precipcover), col="red")
```



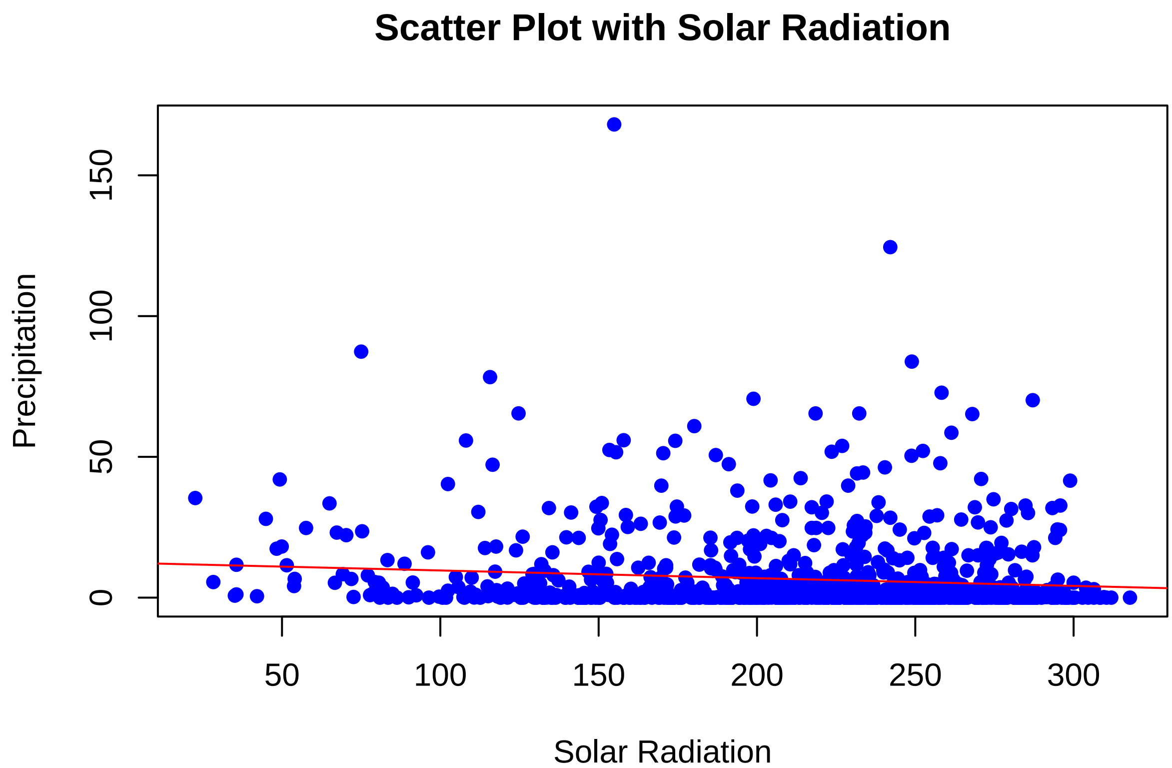

```r
plot(Rainfall$cloudcover, Rainfall$precip,
     main="Scatter Plot with Cloud Cover", xlab="Cloud Cover",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$cloudcover), col="red")
```



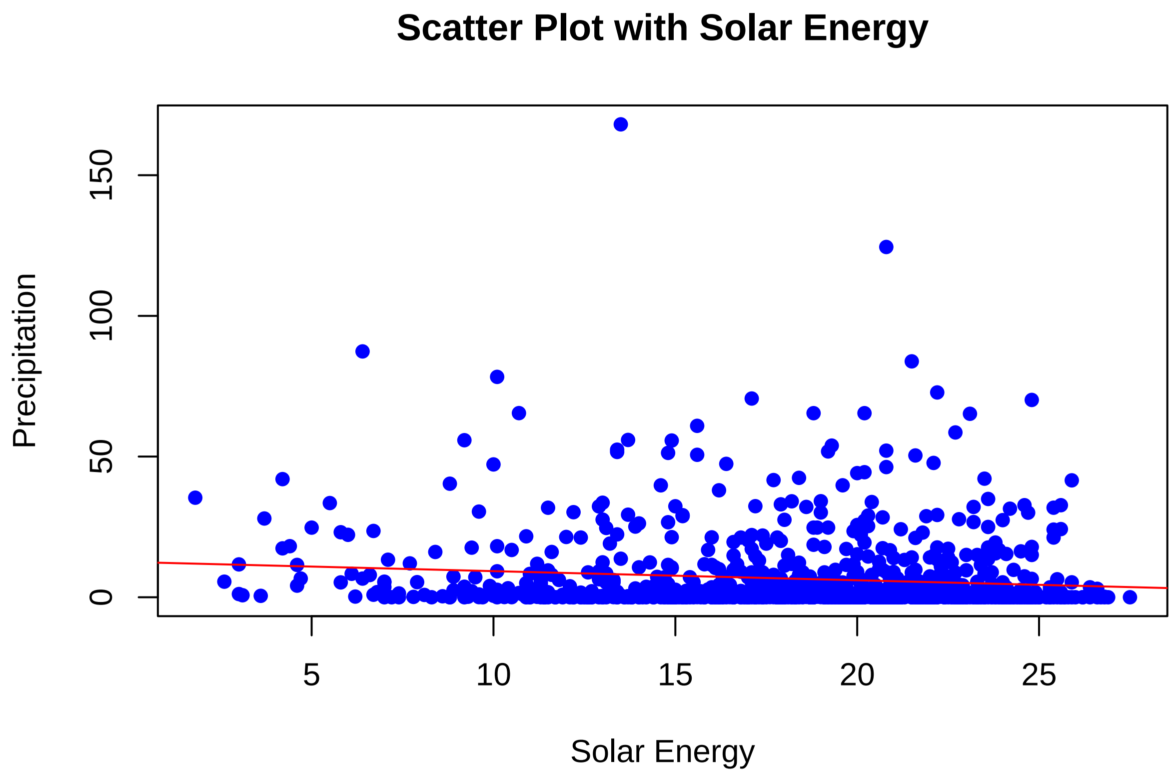

```
plot(Rainfall$visibility, Rainfall$precip,
     main="Scatter Plot with Visibility", xlab="Visibility",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$visibility), col="red")
```




```r
plot(Rainfall$solarradiation, Rainfall$precip,
     main="Scatter Plot with Solar Radiation", xlab="Solar Radiation",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$solarradiation), col="red")
```




```r
plot(Rainfall$solarenergy, Rainfall$precip,
     main="Scatter Plot with Solar Energy", xlab="Solar Energy",
     ylab="Precipitation", col="blue", pch=16)
abline(lm(Rainfall$precip ~ Rainfall$solarenergy), col="red")
```




```
corr<-aov(precip ~ description * uvindex, data=Rainfall)
summary(corr)
```

```
##                      Df Sum Sq Mean Sq F value Pr(>F)
## description           8  39139    4892  30.606 <2e-16 ***
## uvindex               9    825      92   0.573  0.820
## description:uvindex  51   7079     139   0.868  0.732
## Residuals           931 148822     160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

0.6 Multivariate Analysis

##Multiple Linear Regression

```
Precipitation <- lm(precip ~ description + uvindex + tempmax + tempmin +
                      temp + feelslikemax + feelslikemin + feelslike + dew + humidity +
                      precipcover + cloudcover + visibility + solarradiation + solarenergy,
                      data = Rainfall)
anova(Precipitation)
```

```
## Analysis of Variance Table
##
## Response: precip
```

```
##                 Df Sum Sq Mean Sq F value    Pr(>F)    
## description      8  39139  4892.4 37.0930 < 2.2e-16 ***
## uvindex          9    825    91.6  0.6948 0.7140506    
## tempmax          1   1228  1227.9  9.3094 0.0023419 ** 
## tempmin          1   4485  4484.9 34.0034 7.490e-09 ***
## temp             1   5891  5890.6 44.6615 3.946e-11 ***
## feelslikemax     1    577   577.1  4.3756 0.0367174 *  
## feelslikemin     1   1364  1363.8 10.3399 0.0013450 ** 
## feelslike        1    117   116.9  0.8861 0.3467760    
## dew              1   1795  1795.3 13.6115 0.0002373 ***
## humidity         1   9948  9948.1 75.4242 < 2.2e-16 ***
## precipcover      1      7     6.6  0.0501 0.8229131    
## cloudcover       1    475   475.4  3.6043 0.0579260 .  
## visibility       1   1245  1244.6  9.4359 0.0021871 ** 
## solarradiation   1    834   833.6  6.3203 0.0120976 *  
## solarenergy      1    130   130.2  0.9871 0.3207109    
## Residuals      969 127807   131.9                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Precipitation)
```

```
## 
## Call:
## lm(formula = precip ~ description + uvindex + tempmax + tempmin + 
##     temp + feelslikemax + feelslikemin + feelslike + dew + humidity + 
##     precipcover + cloudcover + visibility + solarradiation + 
##     solarenergy, data = Rainfall)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -22.221  -6.111  -1.080   3.443 118.419 
## 
## Coefficients:
##                                                                          Estimate
## (Intercept)                                                            -5.343e+02
## descriptionPartly cloudy throughout the day with afternoon rain.       -1.733e+00
## descriptionPartly cloudy throughout the day with early morning rain.   -4.493e+00
## descriptionPartly cloudy throughout the day with late afternoon rain.   1.162e-01
## descriptionPartly cloudy throughout the day with morning rain.         -1.002e-01
## descriptionPartly cloudy throughout the day with rain clearing later.   2.096e+00
## descriptionPartly cloudy throughout the day with rain in the morning and afternoon.  1.756e+00
## descriptionPartly cloudy throughout the day with rain.                  4.739e+00
## descriptionPartly cloudy throughout the day.                           -4.368e+00
## uvindex2                                                               -3.139e+00
## uvindex3                                                                4.767e-01
## uvindex4                                                               -1.701e+00
## uvindex5                                                               -2.016e+00
## uvindex6                                                               -2.708e+00
## uvindex7                                                               -5.896e+00
## uvindex8                                                               -6.986e+00
## uvindex9                                                               -8.733e+00
## uvindex10                                                              -8.659e+00
## tempmax                                                                 7.294e-01
```

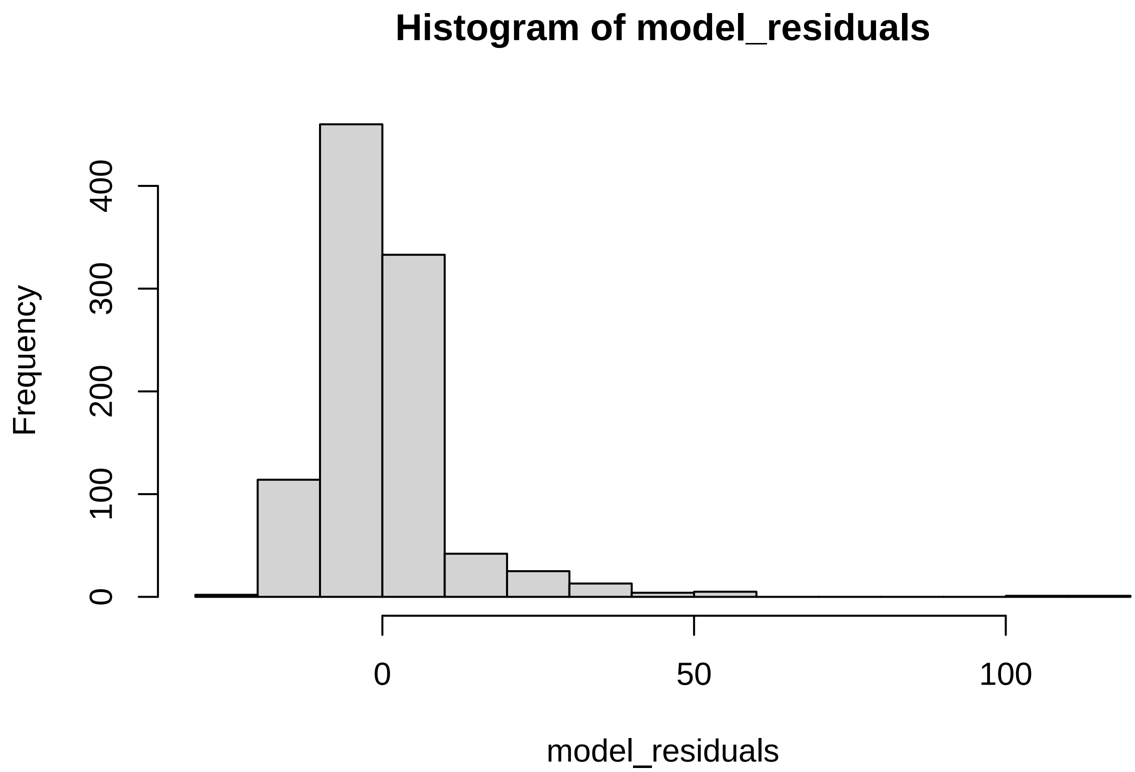
```
## tempmin                                                                     -1.468e+00
## temp                                                                         2.218e+01
## feelslikemax                                                                -7.237e-03
## feelslikemin                                                                 9.085e-01
## feelslike                                                                    3.861e+00
## dew                                                                         -3.351e+01
## humidity                                                                     7.671e+00
## precipcover                                                                  1.542e-02
## cloudcover                                                                   1.130e-01
## visibility                                                                  -2.489e+00
## solarradiation                                                               3.021e-01
## solarenergy                                                                 -3.081e+00
##                                                                             Std. Error
## (Intercept)                                                                  9.840e+01
## descriptionPartly cloudy throughout the day with afternoon rain.             5.758e+00
## descriptionPartly cloudy throughout the day with early morning rain.         5.659e+00
## descriptionPartly cloudy throughout the day with late afternoon rain.        5.576e+00
## descriptionPartly cloudy throughout the day with morning rain.               5.591e+00
## descriptionPartly cloudy throughout the day with rain clearing later.        5.647e+00
## descriptionPartly cloudy throughout the day with rain in the morning and afternoon.  6.043e+00
## descriptionPartly cloudy throughout the day with rain.                       5.174e+00
## descriptionPartly cloudy throughout the day.                                 5.912e+00
## uvindex2                                                                     7.918e+00
## uvindex3                                                                     7.393e+00
## uvindex4                                                                     7.128e+00
## uvindex5                                                                     7.105e+00
## uvindex6                                                                     7.166e+00
## uvindex7                                                                     7.192e+00
## uvindex8                                                                     7.290e+00
## uvindex9                                                                     7.431e+00
## uvindex10                                                                    7.590e+00
## tempmax                                                                      8.323e-01
## tempmin                                                                      1.195e+00
## temp                                                                         4.333e+00
## feelslikemax                                                                 5.061e-01
## feelslikemin                                                                 6.701e-01
## feelslike                                                                    1.291e+00
## dew                                                                          5.015e+00
## humidity                                                                     1.060e+00
## precipcover                                                                  1.324e-01
## cloudcover                                                                   7.472e-02
## visibility                                                                   8.085e-01
## solarradiation                                                               2.681e-01
## solarenergy                                                                  3.101e+00
##                                                                             t value
## (Intercept)                                                                  -5.430
## descriptionPartly cloudy throughout the day with afternoon rain.             -0.301
## descriptionPartly cloudy throughout the day with early morning rain.         -0.794
## descriptionPartly cloudy throughout the day with late afternoon rain.         0.021
## descriptionPartly cloudy throughout the day with morning rain.               -0.018
## descriptionPartly cloudy throughout the day with rain clearing later.         0.371
## descriptionPartly cloudy throughout the day with rain in the morning and afternoon.   0.291
## descriptionPartly cloudy throughout the day with rain.                        0.916
## descriptionPartly cloudy throughout the day.                                 -0.739
```

```
## uvindex2                                                                    -0.396
## uvindex3                                                                     0.064
## uvindex4                                                                    -0.239
## uvindex5                                                                    -0.284
## uvindex6                                                                    -0.378
## uvindex7                                                                    -0.820
## uvindex8                                                                    -0.958
## uvindex9                                                                    -1.175
## uvindex10                                                                   -1.141
## tempmax                                                                      0.876
## tempmin                                                                     -1.229
## temp                                                                         5.118
## feelslikemax                                                                -0.014
## feelslikemin                                                                 1.356
## feelslike                                                                    2.990
## dew                                                                         -6.683
## humidity                                                                     7.239
## precipcover                                                                  0.116
## cloudcover                                                                   1.513
## visibility                                                                  -3.079
## solarradiation                                                               1.127
## solarenergy                                                                 -0.994
##                                                                           Pr(>|t|)
## (Intercept)                                                               7.13e-08
## descriptionPartly cloudy throughout the day with afternoon rain.           0.76348
## descriptionPartly cloudy throughout the day with early morning rain.       0.42736
## descriptionPartly cloudy throughout the day with late afternoon rain.      0.98338
## descriptionPartly cloudy throughout the day with morning rain.             0.98570
## descriptionPartly cloudy throughout the day with rain clearing later.      0.71062
## descriptionPartly cloudy throughout the day with rain in the morning and afternoon.  0.77140
## descriptionPartly cloudy throughout the day with rain.                     0.35993
## descriptionPartly cloudy throughout the day.                               0.46022
## uvindex2                                                                   0.69187
## uvindex3                                                                   0.94860
## uvindex4                                                                   0.81139
## uvindex5                                                                   0.77667
## uvindex6                                                                   0.70555
## uvindex7                                                                   0.41252
## uvindex8                                                                   0.33813
## uvindex9                                                                   0.24016
## uvindex10                                                                  0.25420
## tempmax                                                                    0.38106
## tempmin                                                                    0.21950
## temp                                                                      3.73e-07
## feelslikemax                                                               0.98859
## feelslikemin                                                               0.17550
## feelslike                                                                  0.00286
## dew                                                                       3.95e-11
## humidity                                                                  9.20e-13
## precipcover                                                                0.90731
## cloudcover                                                                 0.13061
## visibility                                                                 0.00214
## solarradiation                                                             0.26003
## solarenergy                                                                0.32071
```

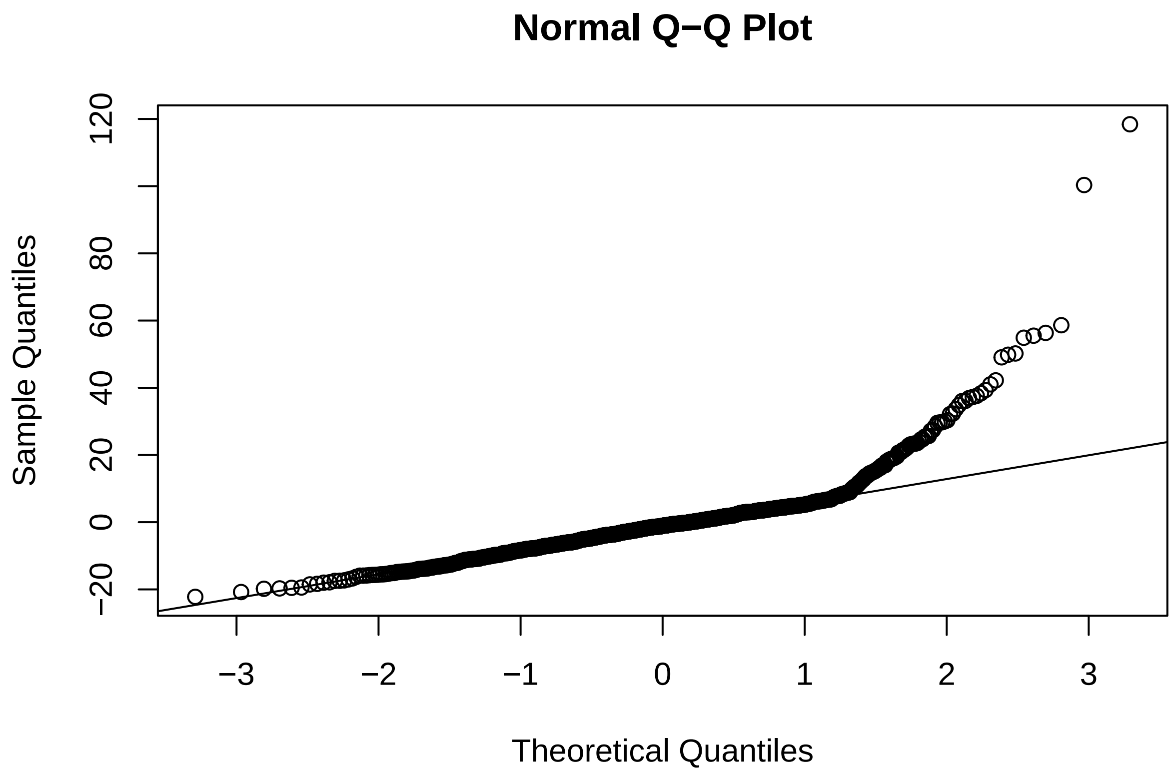
```
##
## (Intercept)                                                                                    ***
## descriptionPartly cloudy throughout the day with afternoon rain.
## descriptionPartly cloudy throughout the day with early morning rain.
## descriptionPartly cloudy throughout the day with late afternoon rain.
## descriptionPartly cloudy throughout the day with morning rain.
## descriptionPartly cloudy throughout the day with rain clearing later.
## descriptionPartly cloudy throughout the day with rain in the morning and afternoon.
## descriptionPartly cloudy throughout the day with rain.
## descriptionPartly cloudy throughout the day.
## uvindex2
## uvindex3
## uvindex4
## uvindex5
## uvindex6
## uvindex7
## uvindex8
## uvindex9
## uvindex10
## tempmax
## tempmin
## temp                                                                                           ***
## feelslikemax
## feelslikemin
## feelslike                                                                                      **
## dew                                                                                            ***
## humidity                                                                                       ***
## precipcover
## cloudcover
## visibility                                                                                     **
## solarradiation
## solarenergy
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.48 on 969 degrees of freedom
## Multiple R-squared:  0.3475, Adjusted R-squared:  0.3273
## F-statistic:  17.2 on 30 and 969 DF,  p-value: < 2.2e-16
```

```r
# Get the model residuals
model_residuals = Precipitation$residuals

# Plot the result
hist(model_residuals)
```




```
# Plot the residuals
qqnorm(model_residuals)
# Plot the Q-Q line
qqline(model_residuals)
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
120
80
60
40
20
0
−20
−3
−2
−1
0
1
2
3