Documentation for Weather Prediction

Using R

**Table of Contents**

[1. Introduction 3](#_Toc10654)

[2. Software Used 3](#_Toc28097)

[3. About R 3](#_Toc12115)

[4. About R Studio 4](#_Toc17760)

[5. Approach 4](#_Toc9436)

[6. List of Libraries Used 4](#_Toc19370)

[7. Packages Installation 5](#_Toc18667)

[8. Process Flow 5](#_Toc30680)

[9. Model Used 5](#_Toc17398)

[10. About ARIMA Model 6](#_Toc11807)

[11. How to Run the Program 6](#_Toc14476)

[12. Program Execution 6](#_Toc24891)

[13. Test Cases 7](#_Toc26429)

[14. Output 8](#_Toc3381)

[15. Source Details 10](#_Toc2752)

# Introduction

* The purpose of this analysis is to find the Weather forecasting based on the given historical information.
* The historical information is taken form web by using the custom package.
* The Output will be shown in the graphical representation for easy interpretation.

# Software Used

The below software are used for the developing this prediction.

Both the software are free Open source tools. These software can be easily downloaded and installed in the local machine.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No** | **Software Name** | **Version** | **Download Location** |
| 1 | R | 3.3 | https://www.r-project.org/ |
| 2 | R Studio Desktop | 1.0.143 | https://www.rstudio.com/products/rstudio/download2/ |

# About R

R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible, R provides an Open Source route to participation in that activity.

One of R’s strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formula where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.

**The R environment**

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

* an effective data handling and storage facility,
* a suite of operators for calculations on arrays, in particular matrices,
* a large, coherent, integrated collection of intermediate tools for data analysis,
* graphical facilities for data analysis and display either on-screen or on hardcopy, can be exported as Image files.
* a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

**<https://www.r-project.org/about.html>**

# About R Studio

RStudio Desktop is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management and does have the Open Source Edition.

# Approach

To take the recent data, predict and check it with the results. So the data is taken from May 1st - May 10th 2017 and project for 10 days, verify the same for the May 11th, 12th

The WeatherData package provides the Daily summarized information with the below set of fields by passing the city code (ie **Airport code**, start date & end date), out of these, I have taken the below 2 fields for the forecasting purpose.

**Available Fields** : Date, Max\_TemperatureC, Mean\_TemperatureC, Min\_TemperatureC, Max\_Sea\_Level\_PressurehPa, Mean\_Sea\_Level\_PressurehPa, Min\_Sea\_Level\_PressurehPa, Precipitationmm, Events

* Maximum Temperature
* Percipitation in MM (Rainfall)

The forecast plotting is put into 2 different graphs to show the Max Temperature & Rainfall separately.

**Note** :

* Since the WeatherData is not providing the location details, latitude, longitude, sea level these are taken from the Net and appended as the part of the data frame. Once the data is pulled the data will be stored in the form of CSV file, in the working directory, for the future purpose. Out file name : **SYDNEY\_Det\_11052017\_1515.csv**
* To do it, the reusable component was written to get the data from web using weatherData package, to check for the null values, save in the flat csv file.

# List of Libraries Used

|  |  |
| --- | --- |
| **Library Name** | **Description** |
| WeatherData | To get the weather data from Web |
| data.table | It offers fast and nemory efficient: file reader and writer, aggregations, updates, equi, non-equi, rolling, range and interval joins |
| lattice | A powerful and elegant high-level data visualization |
| class | Various functions for classification |
| ggplot | To create Elegant data Visualisations |
| forecast | Forecasting Functions for Time Series and Linear Models |
| xts | Extensible Time Series |
| zoo | Infrastructure for Regular and irregular time series |
| funggcast | This is a specific function to convert the ARIMA forecast output to Dataframe, in-turn can be used with ggplot for graphical Representations. (Downloaded from net) |

# Packages Installation

The said packages need to be installed by using the below command in the RStudio. The Package will be pulled from the R CRAN and the installation will be done automatically along-with the installation of the dependencies packages.

**install.packages("weatherData")**

**install.packages("plotrix")**

**install.packages("ggplot2")**

**install.packages("lattice")**

**install.packages("data.table")**

**install.packages("xts”)**

**install.packages("zoo")**

# Process Flow

R Script

Get Weather Data From Web

Apply Model

Do Prediction

Plot the Details in Graph

Graph Export

Common Functions

Weather Prediction

# Model Used

ARIMA model is used for the prediction purpose.

# About ARIMA Model

**ARIMA**stands for Auto Regressive Integrated Moving Average models. Univariate (single vector) **ARIMA**is a forecasting technique that projects the future values of a series based entirely on its own inertia.

# How to Run the Program

* From Command Prompt

Run --> Cmd --> Go to installed Directory

* + **cd C:\Program Files\R\R-3.3.3\bin**

Run the script file using the below command (the scripts are saved in the file)

RScript <path>/**test.R**

**Output**



* **From R Console**

The scripts can be run directly from the R Console. The output will be displayed directly in the console.

* **From R Studio**

R Studio Comes with 4 parts

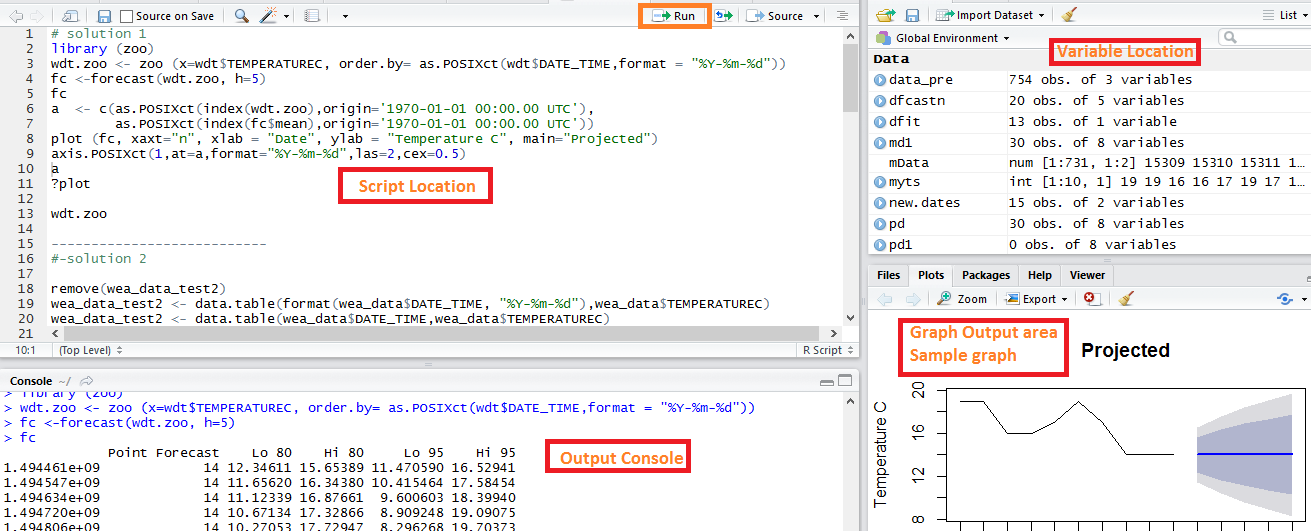
- Script Area

- Output Console

- Variables information Area

- Plot Area

Sample snap shot is given below for easy reference.



# Program Execution

* Open the R Studio
* File --> New File --> R Script / or Press CTRL+SHIFT+N to open the R Script Window
* Set the working directory by using the below command, else the files will be saved in the default directory
  + **setwd**( <required folder>)
  + To know the current directory Run **getwd**() command from the Script window

To Run the **CommonFunctions.R**

* Copy the contents for CommonFunctions.R file, paste it in the Script Window.
* Select all the Lines and Run using the Run button which is available in the right side corner of Script window.
* Check for any errors, if any available, run the scripts individually

To Run the **WeatherPrediction.R**

* Copy the contents for **WeatherPrediction.R** file, paste it in the Script Window.
* Since the Weather Prediction is having the graphs, its advised to run through the R Studio.
* Change the parameters for the given below variables in the console with the required values

city\_cd="SYD" #City code

st\_date="2017-05-01" #Starting date (YYYY-MM-DD)

end\_date="2017-05-10" #Ending date (YYYY-MM-DD)

city\_name="SYDNEY" #City\_name

latlongsl="-33.947346,151.179428,10" #Latitude, longitude, sea level (m)

The code details are available in the attached file



* Run the each and every step.
* There will be 2 portions for graph generation
  + One is for Temperature prediction
  + Second is for Rainfall prediction
* Once the Graph is generated, the graph can be exported by using the export option available in the Graph Area.
* Or the command is available in the script file to export the plot into PDF (the file will be saved in the working directory.
  + The sample file name :
    - For Temperature Prediction : SYDNEY\_FCTemp\_13052017\_1626.pdf
    - For Rainfall Prediction : SYDNEY\_FCRain\_13052017\_1626.pdf

# Test Cases

The Test cases are available in the attached excel sheet.

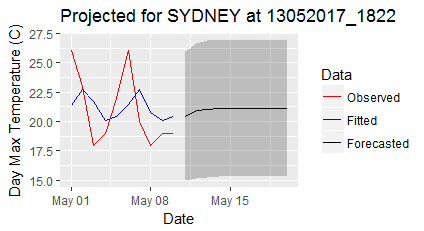


# Output

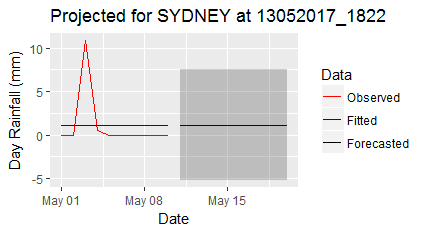
The data is taken for two Cities (Sydney & Melbourne) from May 1st to May 10th 2017, projected for 10 days. The data is analyzed, compared against data for May 11th, 12th and the exact data is falling inside the range of prediction.

**For Sydney**

|  |
| --- |
| **Temperature Predicted** |
| date observed fitted forecast lo80 hi80 lo95 hi95  2017-05-11 NA NA **20.44206** 16.85615 24.02797 14.95788 25.92623  2017-05-12 NA NA **20.90820** 17.13960 24.67679 15.14462 26.67177 |
| **Temperature Recorded** |
| DATE\_TIME MAX\_TEMPERATUREC RAINFALLMM  2017-05-11 **19** 0.00  2017-05-12 **19** 0.51 |

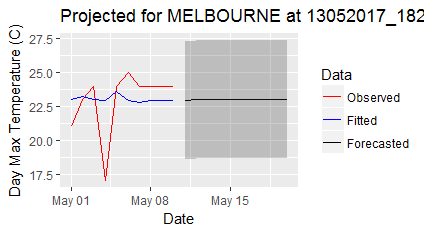


|  |
| --- |
| **Rainfall Predicted** |
| date observed fitted forecast lo80 hi80 lo95 hi95  2017-05-11 NA NA **1.143**  -3.03812 5.32412 -5.251471 7.537471  2017-05-12 NA NA **1.143** -3.03812 5.32412 -5.251471 7.537471 |
| **Rain Fall Recorded** |
| DATE\_TIME MAX\_TEMPERATUREC RAINFALLMM  2017-05-11 19  **0.00**  2017-05-12 19 **0.51** |

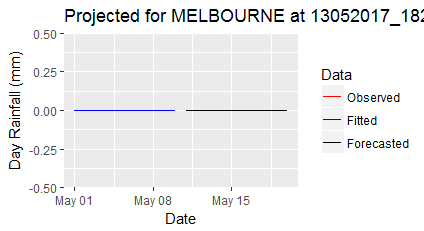


**For Melbourne**

|  |
| --- |
| **Temperature Predicted** |
| date observed fitted forecast lo80 hi80 lo95 hi95  2017-05-11 NA NA **22.90999** 20.06020 25.75979 18.55161 27.26838  2017-05-12 NA NA **23.01923** 20.15516 25.88330 18.63901 27.39945 |
| **Temperature Recorded** |
| DATE\_TIME MAX\_TEMPERATUREC RAINFALLMM  2017-05-11 **26** 0  2017-05-12 **26** 0 |



|  |
| --- |
| **Rainfall Predicted** |
| date observed fitted forecast lo80 hi80 lo95 hi95  2017-05-11 NA NA **0** 0 0 0 0  2017-05-12 NA NA **0** 0 0 0 0 |
| **Rain Fall Recorded** |
| DATE\_TIME MAX\_TEMPERATUREC RAINFALLMM  2017-05-11 26 **0**  2017-05-12 26 **0** |



# Source Details

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **File Name** | **Details about the file** |
| 1 | CommonFunctions.R | This file contains the commonly used R Functions for weather prediction |
| 2 | WeatherPrediction.R | This file is used to predict the weather for Given Location |
| 3 | SYDNEY\_Sum\_13052017\_1449\_before.csv | Weather data in CSV format before prediction |
| 4 | SYDNEY\_Sum\_13052017\_1457\_after.csv | Weather data in CSV format After prediction |
| 5 | MELBOURNE\_Sum\_13052017\_1508\_before.csv | Weather data in CSV format before prediction |
| 6 | MELBOURNE\_Sum\_13052017\_1521\_after.csv | Weather data in CSV format After prediction |
| 7 | SYDNEY\_FCTemp\_13052017\_1825.pdf | Temperature Forecast Graph, generated and saved by the script |
| 8 | SYDNEY\_FCRain\_13052017\_1827.pdf | Rain Forecast Graph, generated and saved by the script |
| 9 | MELBOURNE\_FCTemp\_13052017\_1831.pdf | Temperature Forecast Graph, generated and saved by the script |
| 10 | MELBOURNE\_FCRain\_13052017\_1832.pdf | Rain Forecast Graph, generated and saved by the script. |