



Module Name	Programming for Database
Module Code	CT127-3-2-PFDA
Lecturer Name	Mrs. Minnu Helen Joseph
Hand in Date	1 <sup>st</sup> of April 2021
Hand out date	28 <sup>th</sup> of May 2021
Student ID	TP058435
Student name	Tresor Kalembo Tshibangu

# I. Table of Contents

I. Table of Contents .....	2
I. Introduction.....	3
II. Technical Introduction .....	3
1. Installation and loading of packages .....	3
2. Data import .....	4
3. Data Reading and viewing .....	5
4. Summary of the weather database .....	6
5. Pre-processing.....	7
III. Data analysis .....	9
1. Data Manipulation .....	9
2. Data Visualization.....	<b>Error! Bookmark not defined.</b>
3. Data Exploration .....	<b>Error! Bookmark not defined.</b>
4. Data Manipulation .....	<b>Error! Bookmark not defined.</b>
II. References .....	9

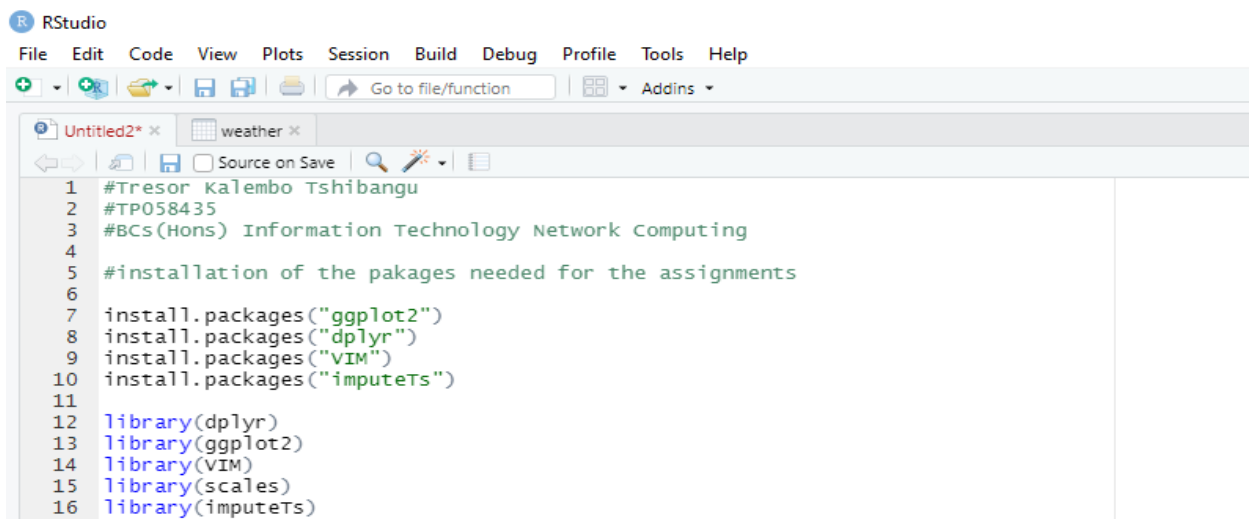
## I. Introduction

In this report, the examination of dataset was done in subtleties and it went farther than what has shrouded in the course. I needed to adjust to the information Import, Visualization, Exploration, and Manipulation ideas which guided me through the arrangement interaction. The motivation behind why I chose to utilize such strategies was additionally advocated.

Some data was pre-handled to get it into a proper organization. data has additionally been defended effectively with proper charts. Additionally, supporting archives that are expected to mirror the diagram and code utilizing R languages ideas have been referenced in this record.

## II. Technical Introduction

### 1. Installation and loading of packages

A screenshot of the RStudio application window. The title bar shows 'RStudio'. The menu bar includes 'File', 'Edit', 'Code', 'View', 'Plots', 'Session', 'Build', 'Debug', 'Profile', 'Tools', and 'Help'. The toolbar contains icons for file operations and a search bar. The main editor window has two tabs: 'Untitled2\*' and 'weather'. The 'Untitled2\*' tab is active, showing R code. The code includes comments for the user's name, ID, and course, followed by instructions for installing and loading several R packages: ggplot2, dplyr, VIM, scales, and imputeTs. The code is as follows:

```
1 #Tresor Kalembo Tshibangu
2 #TP058435
3 #BCs(Hons) Information Technology Network Computing
4
5 #installation of the pakages needed for the assignments
6
7 install.packages("ggplot2")
8 install.packages("dplyr")
9 install.packages("VIM")
10 install.packages("imputeTs")
11
12 library(dplyr)
13 library(ggplot2)
14 library(VIM)
15 library(scales)
16 library(imputeTs)
```

*install.packages* is used to install a package and *library* is used to load the package after its installation. Packages are assortments of R functions, data, and compiled code in a clear-cut way. The index where packages are put away is known as the library. R accompanies a standard arrangement of packages. Others are accessible for download and installation. Once **installed**, they must be load into the session by the use of the function *library* to be utilized.

- *ggplot2* will be utilized for data visualization. The explanation for utilizing is on the grounds that it is known to be a plotting package that simplifies the making of complex plots from data into a data frame. It gives a more automatic interface to determining what factors to plot, how they are shown, and general visual properties. (R Programming Tutorial - Learn the Basics of Statistical Computing, 2019)

- *dplyr* will be utilized for data manipulation and the purpose for utilizing is the need for Speed. Subsequent to exploring I reasoned that *dplyr* is a lot quicker than other, more customary, capacities. Direct association with and investigation inside outer databases allowing less complex treatment of huge data. Function chaining that permits us to try not to mess our workspace with interval objects. (R Programming Tutorial - Learn the Basics of Statistical Computing, 2019)
- *VIM* will be utilized for data pre- processing. The explanation for utilizing *VIM* is on the grounds that the Use *Vim-R-plugin* for altering R code records, R documentation records (\*. Rd) just as the Sweave records. Aside from grammar featuring the plugin permits to open a R console in a different window and work it with console alternate ways from *VIM*. (R Programming Tutorial - Learn the Basics of Statistical Computing, 2019)
- *imputeTS* for Data Pre-Processing. The explanation for utilizing *imputeTS* is on the grounds that it assists me with my missing information issues. The capacities likewise function admirably in tidy style pipe work processes. (R Programming Tutorial - Learn the Basics of Statistical Computing, 2019)
- *mice* (Multivariate Imputation via Chained Equations) is one of the usually utilized package by R clients. Making different imputations when contrasted with a solitary imputation (like mean) deals with vulnerability in values that are missing. *mice* accept that the missing information are Missing at Random (MAR), which implies that the likelihood that a variable is missing relies just upon noticed esteem and can be anticipated utilizing them. It credits data on a variable by factor premise by determining an imputation model for each factor. (R Programming Tutorial - Learn the Basics of Statistical Computing, 2019)
- Stacking the package *scale* will empower me to utilize the *scale* function to scale and focus the data set values involved by the columns of a numeric matrix.

## 2. Data import

We utilize the line of code below to import the dataset from the folder Assignment that contains the database weather which was put there once it was downloaded. It also has a covering capacity for read.table() that commands a comma as separator and utilizations the input document's first line as header that determines the column names of the table. Consequently, it is an ideal contender to peruse CSV documents.

```

18
19 #import weather database and assigned it to the named weather
20
21 weather <- read.csv("C:/Users/HP/Desktop/Level 2 Assignments/PDFA Assignment/Assignment/weather.csv")
22

```

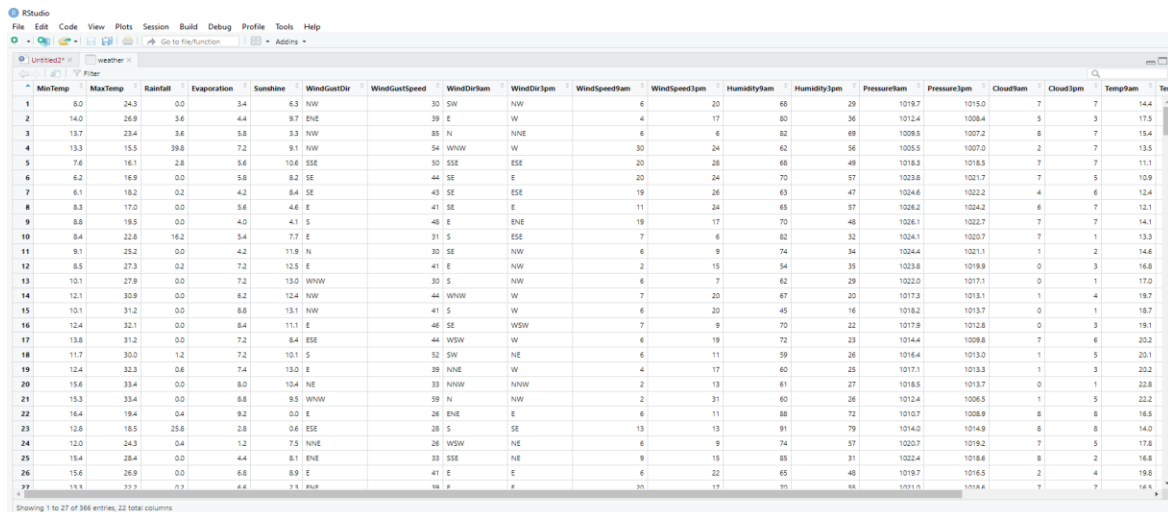
### 3. Data Reading and viewing

```
22
23 #viewing and printing the weather database
24
25 view(weather)
26 print(weather)
27
```

To ensure that the *weather* is appropriately imported, we need to utilize the above code to view the all database on another tab just as print the dataset to confirm it is stacking appropriately in the terminal.

with the use of *view* function we can see that another tab is opened in the R-Studio which gives a full perspective on the database *weather*, after using *print* function we can see that the few values is printed, which is not the full database, on to the console for us to read.

Output for *view*:



	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDirAm	WindDirSpn	WindSpeedAm	WindSpeedSpn	HumidityAm	HumiditySpn	PressureAm	PressureSpn	CloudAm	CloudSpn	TempAm	TempSpn
1	8.0	24.3	0.0	3.4	6.3	NW		30	SW	NW	6	20	68	29	1019.7	1015.0	7	7	14.4
2	14.0	26.9	3.6	4.4	9.7	ENE		39	E	W	4	17	80	36	1012.4	1008.4	5	3	17.5
3	13.7	23.4	3.6	5.8	3.3	NW		85	N	NNE	6	6	82	69	1009.5	1007.2	8	7	13.4
4	13.3	15.5	39.8	7.2	9.1	NW		54	WNW	W	30	24	62	56	1005.5	1007.0	2	7	13.5
5	7.6	16.1	2.6	5.6	10.6	SSE		30	SSE	ESE	20	28	68	49	1018.3	1018.5	7	7	11.1
6	6.2	16.9	0.0	5.8	8.2	SE		44	SE	E	20	24	70	57	1023.8	1021.7	7	5	10.9
7	6.1	18.2	0.2	4.2	8.4	SE		43	SE	ESE	19	26	63	47	1024.6	1022.2	4	6	12.4
8	8.3	17.0	0.0	5.6	4.6	E		41	SE	E	11	24	65	57	1026.2	1024.2	6	7	12.1
9	8.8	19.5	0.0	4.0	4.1	S		48	E	ENE	19	17	70	48	1026.1	1022.7	7	7	14.1
10	8.4	22.6	16.2	5.4	7.7	E		31	S	ESE	7	6	82	32	1024.1	1020.7	7	1	13.3
11	9.1	25.2	0.0	4.2	11.9	N		30	SE	NW	6	9	74	34	1024.4	1021.1	1	2	14.6
12	8.5	27.3	0.2	7.2	12.3	E		41	E	NW	2	15	54	35	1023.8	1019.9	0	3	16.8
13	10.1	27.9	0.0	7.2	13.0	WNW		30	S	NW	6	7	62	29	1022.0	1017.1	0	1	17.0
14	12.1	30.9	0.0	6.2	12.4	NW		44	WNW	W	7	20	67	20	1017.3	1018.1	1	4	18.7
15	10.1	31.2	0.0	8.8	13.1	NW		41	S	W	6	20	45	16	1018.2	1013.7	0	1	18.7
16	12.4	32.1	0.0	8.4	11.1	E		46	SE	WSW	7	9	70	22	1017.9	1012.8	0	3	19.1
17	13.8	31.2	0.0	7.2	8.4	ESE		44	WSW	W	6	19	72	23	1014.4	1009.8	7	6	20.2
18	11.7	30.0	1.2	7.2	10.1	S		52	SW	NE	6	11	59	26	1016.4	1013.0	1	5	20.1
19	12.4	32.3	0.6	7.4	13.0	E		39	NNE	W	4	17	60	25	1017.1	1013.3	1	3	20.2
20	15.6	33.4	0.0	8.0	10.4	NE		33	NNW	NNW	2	13	61	27	1018.5	1013.7	0	1	22.8
21	15.3	33.4	0.0	8.0	9.5	WNW		59	N	NW	2	31	60	26	1012.4	1006.5	1	5	22.2
22	16.4	19.4	0.4	9.2	0.0	E		26	ENE	E	6	11	88	72	1010.7	1008.9	8	8	16.5
23	12.8	18.5	25.8	2.8	0.6	ESE		28	S	SE	13	13	91	79	1014.0	1014.9	8	8	14.0
24	12.0	24.3	0.4	1.2	7.5	NNE		26	WSW	NE	6	9	74	57	1020.7	1019.2	7	5	17.8
25	18.4	28.4	0.0	4.4	8.1	ENE		33	SSE	NE	9	15	85	31	1022.4	1018.8	8	2	16.8
26	15.6	26.9	0.0	6.8	8.9	E		41	E	E	6	22	65	48	1018.7	1016.5	2	4	19.8
27	13.3	27.3	0.1	6.6	7.1	NE		18	E	E	21	17	75	45	1011.5	1018.4	7	7	18.4

Output for *print*:

```

> View(weather)
> print(weather)
  MIntTemp MaxTemp Rainfall Evaporation Sunshine windgustDir windgustSpeed windDir9am windDir3pm windSpeed9am windSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm
1      8.0    21.3      0.0      3.4      6.3      NW      30      SW      6      20      68      29      1019.7      1015.0      7      7      14.4      23.6
2     14.0    26.9      3.6      4.4      9.7      ENE      39      E      W      4      17      80      36      1012.4      1008.4      5      3      17.5      25.7
3     13.7    23.4      3.6      5.8      3.3      NW      85      N      NNE      6      6      82      69      1009.5      1007.2      8      7      15.4      20.2
4     13.3    15.5     39.8      7.2      9.1      NW      54      WNW      W      30      24      62      56      1005.5      1007.0      2      7      13.5      14.1
5      7.6    16.1      2.8      5.6     10.6      SSE      50      SSE      ESE      20      28      68      49      1018.3      1018.5      7      7      11.1      15.4
6      6.2    16.9      0.0      5.8      8.2      SE      44      SE      E      20      24      70      57      1023.8      1021.7      7      5      10.9      14.8
7      6.1    18.2      0.2      4.2      8.4      SE      43      SE      ESE      19      26      63      47      1024.6      1022.2      4      6      12.4      17.3
8      8.3    17.0      0.0      5.6      4.6      E      41      SE      E      11      24      65      57      1026.2      1024.2      6      7      12.1      15.5
9      8.8    19.5      0.0      4.0      4.1      S      48      E      ENE      19      17      70      48      1026.1      1022.7      7      7      14.1      18.9
10     8.4    22.8     16.2      5.4      7.7      E      31      S      ESE      7      6      82      32      1024.1      1020.7      7      1      13.3      21.7
11     9.1    25.2      0.0      4.2     11.9      N      30      SE      NW      6      9      74      34      1024.4      1021.1      1      2      14.6      24.0
12     8.5    27.3      0.2      7.2     12.5      E      41      E      NW      2      15      54      35      1023.8      1019.9      0      3      16.8      26.0
13    10.1    27.9      0.0      7.2     13.0      WNW      30      S      NW      6      7      62      29      1022.0      1017.1      0      1      17.0      27.1
14    12.1    30.9      0.0      6.5     12.4      NW      44      WNW      W      7      20      67      20      1017.3      1013.1      1      4      19.7      30.7
15    10.1    31.2      0.0      8.8     13.1      NW      41      S      W      6      20      45      16      1018.2      1013.7      0      1      18.7      30.4
16    12.4    32.1      0.0      8.4     11.1      E      46      SE      WSW      7      9      72      23      1014.4      1009.8      7      6      20.2      29.8
17    13.8    31.2      0.0      7.2      8.4      ESE      44      WSW      W      6      19      72      23      1014.4      1009.8      7      6      20.2      29.8
18    11.7    30.0      1.2      7.2     10.1      S      52      SW      NE      6      11      59      26      1016.4      1013.0      1      5      20.1      28.6
19    12.4    32.3      0.6      7.4     13.0      E      39      NNE      W      4      17      60      25      1017.1      1013.3      1      3      20.2      31.2
20    15.6    33.4      0.0      8.0     10.4      NE      33      NNW      NNW      2      13      61      27      1018.5      1013.7      0      1      22.8      32.0
21    15.3    33.4      0.0      8.8      9.5      WNW      59      N      NW      2      31      60      26      1012.4      1006.5      1      5      22.2      32.8
22    16.4    19.4      0.4      9.2      0.0      E      26      ENE      E      6      11      88      72      1010.7      1008.9      8      8      16.5      18.3
23    12.8    18.5     25.8      2.8      0.6      ESE      28      S      SE      13      13      91      79      1014.0      1014.9      8      8      14.0      16.8
24    12.0    24.3      0.4      1.2      7.5      NNE      26      WSW      NE      6      9      74      57      1020.7      1019.2      7      5      17.8      22.8
25    15.4    28.4      0.0      4.4      8.1      ENE      33      SSE      NE      9      15      85      31      1022.4      1018.6      8      2      16.8      27.3
26    15.6    26.9      0.0      6.8      8.9      E      41      E      E      6      22      65      48      1019.7      1016.5      2      4      19.8      25.1
27    13.3    22.2      0.2      6.6      2.3      ENE      39      E      E      20      17      70      55      1021.0      1018.6      7      7      16.5      21.2
28    12.9    28.0      0.0      4.4     10.7      S      52      S      NNE      6      11      61      31      1019.2      1014.8      5      7      18.8      26.7
29    15.1    24.3      0.0      7.0      0.4      SE      39      SE      SE      7      17      69      80      1019.0      1017.1      7      7      18.9      19.7
30    13.6    24.1      0.4      2.6      0.5      NNW      30      SSW      S      6      6      83      49      1017.2      1013.3      8      7      17.3      23.2
31    15.1    20.4     22.6      2.4      0.2      SSE      41      E      S      6      20      82      90      1015.0      1014.1      8      8      17.0      16.3
32    11.6    26.3      4.2      1.6     12.0      NNE      41      E      SW      6      9      66      46      1013.9      1009.5      2      2      16.7      25.2
33    16.6    24.2      0.2      6.6      4.7      NW      50      WNW      NW      13      35      72      60      1007.9      1004.6      7      7      20.5      19.9
34    13.3    26.5      6.6      3.8     11.8      NW      50      NW      WNW      20      26      74      40      1007.3      1006.3      5      2      18.8      25.1
35    14.5    21.8      0.0      8.4      9.8      ENE      43      ESE      E      11      30      69      63      1015.0      1015.3      7      1      18.6      20.5
36    16.3    26.8      0.0      6.0      6.3      ENE      39      ESE      ESE      13      9      78      54      1018.8      1016.1      8      7      18.6      24.5
37    17.2    25.8      0.0      4.2      8.8      SW      41      NW      N      6      6      70      74      1014.5      1011.5      6      7      21.5      22.6
38    16.5    28.2      4.0      4.2      8.8      NE      39      E      N      7      13      76      42      1014.8      1011.4      8      1      18.0      26.3
39    15.0    29.4      0.0      6.6     11.1      NW      43      N      W      9      24      78      24      1010.8      1008.5      7      7      20.6      28.6
40    14.9    24.8      0.0      10.4     10.0      NNW      35      NNE      WNW      11      6      58      37      1011.7      1010.4      7      7      19.3      22.9
41    11.8    18.5      0.6      4.8      2.3      ENE      35      ESE      E      9      15      60      63      1018.0      1017.4      8      6      13.8      16.4
42    11.7    21.5      0.0      4.2      7.3      ENE      41      ESE      E      15      24      66      51      1021.1      1019.6      7      5      15.7      19.4
43      9.6    20.3      0.0      5.0      3.6      SE      39      ESE      E      22      20      70      49      1021.9      1019.3      7      7      15.8      19.2
44      8.9    27.1      0.0      4.4     12.7      E      35      NNW      N      6      11      63      37      1018.1      1012.5      1      2      15.8      25.0
45    10.1    29.9      0.0      6.8      8.8      E      41      SE      WNW      2      13      63      36      1012.6      1010.2      1      8      17.4      29.1

RainToday RISK_MM RainTomorrow
1      No      3.6      Yes
2      Yes      3.6      Yes
3      Yes     39.8      Yes
4      Yes      2.8      Yes
5      Yes      0.0      No
6      No      0.2      No

```

#### 4. Summary of the weather database

```

27
28 #different type of summary to have an idea of the database
29 summary(weather)
30 attributes(weather)
31
32

```

- **summary** is a nonexclusive function used to deliver result synopses of the consequences of different model fitting functions. The function conjures specific techniques which rely upon the class of the 1<sup>st</sup> argument. Its output shows you for each factor a bunch of elucidating measurements, contingent upon the sort of the variable: Numerical factors: **summary** gives you the mean, quartiles, range, and median. (DataCamp, 2021)

- **colnames** function is utilized to set the names to segments of a framework with a syntax: `colnames(x) <-value`.

Parameters:

x: Matrix.

Utilizing the **colnames** function we can see the names of the section in our dataset. These are otherwise called attributes which we will break down later. (geeksforgeeks, 2020)

- Items in R can have numerous properties related with them, called **attributes**. These properties clarify what an article addresses and how it ought to be deciphered by R. Frequently, the solitary contrast between two comparable articles is that they have

various attributes. Utilizing *attribute* function, we can see diverse information about the database weather and its main components. (O'Reilly Media, Inc., 2021)

Output for *attributes(weather)*:

```
> attributes(weather)
$names
[1] "MinTemp" "MaxTemp" "Rainfall" "Evaporation" "Sunshine" "windGustDir" "windGustSpeed" "windDir9am" "windDir3pm" "windSpeed9am"
[11] "windSpeed3pm" "Humidity9am" "Humidity3pm" "Pressure9am" "Pressure3pm" "Cloud9am" "Cloud3pm" "Temp9am" "Temp3pm" "RainToday"
[21] "RISK_MM" "RainTomorrow"

$class
[1] "data.frame"

$row.names
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
[44] 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86
[87] 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129
[130] 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172
[173] 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215
[216] 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258
[259] 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301
[302] 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344
[345] 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366

> |
```

Output for *summary(weather)*:

```
> summary(weather)
MinTemp      Min.      -5.300      1st Qu.: 2.300      Median : 7.450      Mean   : 7.266      3rd Qu.:12.500      Max.    :20.900
MaxTemp      Min.       7.60      1st Qu.:15.03      Median :19.65      Mean   :20.55      3rd Qu.:25.50      Max.    :35.80
Rainfall     Min.       0.000      1st Qu.: 0.000      Median : 0.000      Mean   : 1.428      3rd Qu.: 0.200      Max.    :39.800
Evaporation  Min.       0.200      1st Qu.: 2.200      Median : 4.200      Mean   : 4.522      3rd Qu.: 6.400      Max.    :13.800
Sunshine     Min.       0.000      1st Qu.: 5.950      Median : 8.600      Mean   : 7.909      3rd Qu.:10.500      Max.    :13.600
windGustDir  Length:366      Class :character      Mode :character
windGustSpeed Min.      :13.00      1st Qu.:31.00      Median :39.00      Mean   :39.84      3rd Qu.:46.00      Max.    :98.00
windDir9am   Length:366      Class :character      Mode :character
windDir3pm   Length:366      Class :character      Mode :character
windSpeed9am Min.      : 0.000      1st Qu.: 6.000      Median : 7.000      Mean   : 9.652      3rd Qu.:13.000      Max.    :41.000
windSpeed3pm Min.      :36.00      1st Qu.:64.00      Median :72.00      Mean   :72.04      3rd Qu.:84.00      Max.    :99.00
Humidity9am  Min.      :13.00      1st Qu.:32.25      Median :43.00      Mean   :44.52      3rd Qu.:55.00      Max.    :96.00
Humidity3pm  Min.      :96.5      1st Qu.:101.5      Median :102.0      Mean   :101.7      3rd Qu.:102.5      Max.    :103.7
Pressure9am  Min.      :996.5      1st Qu.:1012.8      Median :1017.4      Mean   :1016.8      3rd Qu.:1021.5      Max.    :1033.2
Pressure3pm  Min.      :996.8      1st Qu.:1012.8      Median :1017.4      Mean   :1016.8      3rd Qu.:1021.5      Max.    :1033.2
Cloud9am     Min.      :0.000      1st Qu.:1.000      Median :3.500      Mean   :3.891      3rd Qu.:7.000      Max.    :8.000
Cloud3pm     Min.      :0.000      1st Qu.:1.000      Median :4.000      Mean   :4.025      3rd Qu.:7.000      Max.    :8.000
Temp9am      Min.      : 0.100      1st Qu.: 7.625      Median :12.550      Mean   :12.358      3rd Qu.:17.000      Max.    :24.700
Temp3pm      Min.      : 5.10      1st Qu.:14.15      Median :18.55      Mean   :19.23      3rd Qu.:24.00      Max.    :34.50
RainToday    Length:366      Class :character      Mode :character
RISK_MM      Min.      : 0.000      1st Qu.: 0.000      Median : 1.428      Mean   : 1.428      3rd Qu.: 0.200      Max.    :39.800
RainTomorrow Length:366      Class :character      Mode :character

> |
```

Output for *colnames(weather)*:

```
> colnames(weather)
[1] "MinTemp" "MaxTemp" "Rainfall" "Evaporation" "Sunshine" "windGustDir" "windGustSpeed" "windDir9am"
[9] "windDir3pm" "windSpeed9am" "Humidity9am" "Humidity3pm" "Pressure9am" "Pressure3pm" "Cloud9am"
[17] "Cloud3pm" "Temp9am" "Temp3pm" "RainToday" "RISK_MM" "RainTomorrow"

> |
```

## 5. Pre-processing and Transformations

The *distinct* is utilized to clean the data set *weather* from each similar data. This is on the grounds that we need our data set to be exceptional to build the speed while searching for specific data. In the wake of cleaning the dataset, it is allocated to another variable named *newDataset* to not influence the first data set so we can have a reference while looking at if there was any similar data. (DataNovia, 2018)

The *deduplicated* returns a steady vector where TRUE shows what parts of a vector or data set are indistinguishable. *!deduplicated* infers that we do not need to be bothered with the copy rows. (DataNovia, 2018)

```
> any(is.na(newDataset))#To check if there is any data that is missing in newDataset
[1] TRUE
> which(is.na(newDataset))#to have an image of how many data that are missing.
[1] 1686 1765 1813 2024 2042 2102 2408 2468 2676 2681 2690 2701 2703 2737 2739 2743 2752 2758 2763 2771 2772 2775 2776 2800 2810 2812 2813 2818
[29] 2820 2822 2825 2835 2836 2841 2862 2865 2875 2914 2923 3151 3408 3413 3422 3490 3508 3542 3552
> newData = na.omit(newDataset)
> newData = na.omit(newData)
> any(is.na(newData))
[1] FALSE
> which(is.na(newData))
integer(0)
> |
```



### III. Data analysis (Manipulation, visualization, and exploration)

#### 1. Question 1: When is the worst time to go to the Park with the kids

Analysis 1: Find the coldest time

Analysis 2: Find the windiest time.

Analysis 3: Find time when it rains less

#### 2. Question 2: When is the worst time to go to the beach

Analysis 1: Find the sunniest time

Analysis 2: Find the less windy time.

Analysis 3: Find the no raining time.

#### 3. Question 3: When is the best time to have a barbecue

Analysis 1: Find the no raining time.

Analysis 2: Find the less windy time.

Analysis 3: Find the sunny time.

#### 4. Question 4: Can I wear a short and a vest?

Analysis 1: Find the no raining time.

Analysis 2: Find the less windy time.

Analysis 3: Find the time when MinTemp is less than 10.

## II. References

DataCamp, 2021. *topics : md.pattern*. [Online]

Available at: <https://www.rdocumentation.org/packages/mice/versions/3.13.0/topics/md.pattern>

[Accessed 28 May 2021].

DataNovia, 2018. *DATA MANIPULATION IN R: IDENTIFY AND REMOVE DUPLICATE*

*DATA IN R*. [Online]

Available at: [https://www.datanovia.com/en/lessons/identify-and-remove-duplicate-data-in-](https://www.datanovia.com/en/lessons/identify-and-remove-duplicate-data-in-r/#:~:text=The%20function%20distinct()%20%5Bdplyr,R%20base%20function%20unique()%20)

[r/#:~:text=The%20function%20distinct\(\)%20%5Bdplyr,R%20base%20function%20unique\(\)%20](https://www.datanovia.com/en/lessons/identify-and-remove-duplicate-data-in-r/#:~:text=The%20function%20distinct()%20%5Bdplyr,R%20base%20function%20unique()%20)

[0.](https://www.datanovia.com/en/lessons/identify-and-remove-duplicate-data-in-r/#:~:text=The%20function%20distinct()%20%5Bdplyr,R%20base%20function%20unique()%20)

[Accessed 28 May 2021].

Kabacoff.PhD, R. I., 2017. *input : missingdata*. [Online]

Available at: <https://www.statmethods.net/input/missingdata.html>

[Accessed 28 May 2021].