Table of Contents

1.0 Programming Technique

1.1 install package

install.packages("ggplot2"), install.packages("dplyr"), install.packages("crayon"), library(crayon), library(ggplot2), library(dplyr)

Code above is the code for installing packages. Packages in R is a collection of R function, complied code and sample data. And they stored in a directory called “library”.

1.2 Read Files

weather = read.csv(file="C:\\Users\\Gohiki\\Desktop\\weather.csv", header=TRUE, sep=",")

Code above is use for reading excel file in R programming.

```
> head(weather)
  MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir WindGustSpeed WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am
1     8.0    24.3      0.0         3.4      6.3          NW            30         SW          NW            6           20           68           29      1019.7
2    14.0    26.9      3.6         4.4      9.7         ENE            39          E           W            4           17           80           36      1012.4
3    13.7    23.4      3.6         5.8      3.3          NW            85          N         NNE            6            6           82           69      1009.5
4    13.3    15.5     39.8         7.2      9.1          NW            54        WNW           W           30           24           62           56      1005.5
5     7.6    16.1      2.8         5.6     10.6         SSE            50        SSE         ESE           20           28           68           49      1018.3
6     6.2    16.9      0.0         5.8      8.2          SE            44         SE           E           20           24           70           57      1023.8
  Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday RISK_MM RainTomorrow
1      1015.0        7        7    14.4    23.6        No     3.6          Yes
2      1008.4        5        3    17.5    25.7       Yes     3.6          Yes
3      1007.2        8        7    15.4    20.2       Yes    39.8          Yes
4      1007.0        2        7    13.5    14.1       Yes     2.8          Yes
5      1018.5        7        7    11.1    15.4       Yes     0.0           No
6      1021.7        7        5    10.9    14.8        No     0.2           No
>
```

Figure 1 weather

Figure above used function head(weather) to view the first 6 variable in the excel file.

1.3 Data Manipulation

Data manipulation is modifying data and make the data easier to read. The package will be used is “dplyr” package. In “dplyr” package, it has select(), filter(), mutate(), summarise(), group_by() and join() function. And the function will be used in this assignment is select(), filter(), mutate(), summarise() and group_by().

Filter() is use for filtering the data which match the criteria and show it out. Select() is select the variable user wants. Mutate() is use for creating new variable with the existing variables. Arrange() is use for ordering of the row. Summarise() is use for reducing value for making summary of the variable. Group_by() is use for changing the scope of each function from operating on the dataset. Also, %>% which is also called pipes, use for chaining several functions together.

1.3.1 Data Manipulation 1

```
# 1 Filter
ContinueRain=filter(weather, RainToday =="Yes" , RainTomorrow == "Yes")
```

Figure 2 manipulation 1

This code is used for filtering the data for raining today and tomorrow. And the data will be named as ContinueRain.

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am
1	14.0	26.9	3.6	4.4	9.7	ENE	39	E
2	13.7	23.4	3.6	5.8	3.3	NW	85	N
3	13.3	15.5	39.8	7.2	9.1	NW	54	WNW
4	15.1	20.4	22.6	2.4	0.2	SSE	41	E
5	15.5	21.1	5.4	6.4	0.9	S	31	SSE
6	12.6	23.1	3.4	1.6	2.3	NNW	30	N

Figure 3 Continue Rain 1

Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
1008.4	5	3	17.5	25.7	Yes	3.6	Yes
1007.2	8	7	15.4	20.2	Yes	39.8	Yes
1007.0	2	7	13.5	14.1	Yes	2.8	Yes
1014.1	8	8	17.0	16.3	Yes	4.2	Yes
1008.6	8	8	16.6	20.0	Yes	1.4	Yes
1014.1	8	7	15.3	20.4	Yes	6.4	Yes

Figure 4 Continue 2

WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am
W	4	17	80	36	1012.4
NNE	6	6	82	69	1009.5
W	30	24	62	56	1005.5
S	6	20	82	90	1015.0
NE	6	7	93	86	1010.1
NW	4	13	97	74	1015.8

Figure 5 Continue Rain 3

Figure above is the output data of the Continue Rain. And it shows the data when today and tomorrow is raining.

1.3.2 Data Manipulation 2

```
NoRain= weather %>% na.omit %>% filter( RainToday =="No" , RainTomorrow == "No")
```

Figure 6 manipulation 2

Code above is to filter when today and tomorrow doesn't have rain. And for the na.omit, it is use for removing NA (not available) data. Also, this code will be named as NoRain.

MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed
6.2	16.9	0.0	5.8	8.2	SE	44
6.1	18.2	0.2	4.2	8.4	SE	43
8.3	17.0	0.0	5.6	4.6	E	41
9.1	25.2	0.0	4.2	11.9	N	30
8.5	27.3	0.2	7.2	12.5	E	41
10.1	27.9	0.0	7.2	13.0	WNW	30

Figure 7 No rain data1

WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am
E	20	24	70	57	1023.8
ESE	19	26	63	47	1024.6
E	11	24	65	57	1026.2
NW	6	9	74	34	1024.4
NW	2	15	54	35	1023.8
NW	6	7	62	29	1022.0

Figure 8 no rain data 2

Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RISK_MM	RainTomorrow
1021.7	7	5	10.9	14.8	No	0.2	No
1022.2	4	6	12.4	17.3	No	0.0	No
1024.2	6	7	12.1	15.5	No	0.0	No
1021.1	1	2	14.6	24.0	No	0.2	No
1019.9	0	3	16.8	26.0	No	0.0	No
1017.1	0	1	17.0	27.1	No	0.0	No

Figure 9 norain data3

Figure above is the data when today and tomorrow is not raining. The data is viewed by using head(). Head() will show the first 6 data of your dataset.

1.3.3 Data Manipulation 3

```
weather%>%summarise(max(MaxTemp))
```

Figure 10 manipulation 3

Code above used summarise(). Summarise() is use for calculating the summary of the selected data mean, max, min and others. In this code, max is used for checking the highest Temperature of MaxTemp data.

```
  max(MaxTemp)
1         35.8
```

Figure 11 maxMaxTemp

As shown above, the highest temperature in MaxTemp is 35.8 degree Celsius.

1.3.4 Data Manipulation 4

```
weather %>% filter(RainToday =="Yes") %>% summarise(max(Evaporation))
```

Figure 12 data manipulation 4

Code above is used for check what is the maximum Evaporation of the dataset when today is raining.

```
max(Evaporation)
            10.2
```

Figure 13 maxEvaporation

Figure above is the output for max(Evaporation) when today is raining. As shown above, the evaporation when today is raining is 10.2.

1.3.5 Data Manipulation 5

```
weather%>%filter(RainToday == "Yes" , RainTomorrow=="Yes") %>% select(MinTemp, MaxTemp, Evaporation, WindGustSpeed, RainToday, RainTomorrow)
```

Figure 14 data manipulation 5

Code above is use for checking the details of MinTemp, MaxTemp, Evaporation, WindGustSpeed, RainToday and RainTomorrow when today and tomorrow is raining.

	MinTemp	MaxTemp	Evaporation	WindGustSpeed	RainToday	RainTomorrow
1	14.0	26.9	4.4	39	Yes	Yes
2	13.7	23.4	5.8	85	Yes	Yes
3	13.3	15.5	7.2	54	Yes	Yes
4	15.1	20.4	2.4	41	Yes	Yes
5	15.5	21.1	6.4	31	Yes	Yes
6	12.6	23.1	1.6	30	Yes	Yes
7	14.8	29.5	1.8	41	Yes	Yes
8	19.9	22.0	4.4	76	Yes	Yes
9	15.3	19.6	5.0	33	Yes	Yes
10	17.6	27.8	10.2	39	Yes	Yes
11	18.2	22.6	8.0	33	Yes	Yes
12	14.5	24.2	6.6	48	Yes	Yes
13	16.3	24.8	7.8	50	Yes	Yes
14	13.1	17.4	2.8	43	Yes	Yes
15	7.1	19.8	3.2	39	Yes	Yes
16	4.7	18.5	2.0	22	Yes	Yes
17	8.6	13.7	2.2	31	Yes	Yes
18	6.1	17.2	2.2	59	Yes	Yes
19	13.1	19.4	8.8	67	Yes	Yes
20	8.7	19.7	5.2	98	Yes	Yes
21	14.4	20.7	9.4	33	Yes	Yes

Figure 15 output manipulation 5

From figure above, the data that got when today and tomorrow is raining is 26. The lowest temperature when today and tomorrow is raining is 4.7, maximum temperature is 26.9.

1.3.6 Data Manipulation 6

```
HighestWGS = weather %>% na.omit %>% group_by(WindGustDir) %>% select(WindGustSpeed, WindGustDir) %>% summarise(max(WindGustSpeed))
```

Figure 16 manipulation 6

Code above is use for finding the highest wind speed for each direction in the dataset. The function used in this code is group_by(), select() and summarise().

	WindGustDir	`max(WindGustSpeed)`
	<fct>	<int>
1	E	50
2	ENE	46
3	ESE	50
4	N	59
5	NE	39
6	NNE	41
7	NNW	83
8	NW	98
9	S	70
10	SE	48
11	SSE	57
12	SSW	70
13	SW	50
14	W	70
15	WNW	65
16	WSW	46

Figure 17 output 6

Figure above shows every direction the wind facing and the highest wind speed of the wind. As we can see, it has 16 direction of the wind facing, which is E, ENE, ESE, N, NE, NNE, NNW, NW, S, SE, SSE, SSW, SW, W, WNW and WSW. And each of the wind direction have the highest wind speed.

1.3.7 Data Manipulation 7

```
WindDirection = weather %>% select(WindGustDir, WindGustSpeed, WindSpeed3pm, WindSpeed9am)
summary(WindDirection)
```

Figure 18 manipulation 7

Figure above is the code for find each of the wind direction that the detail of wind gust speed, wind gust speed on 9 am and 3 pm, and making a summary by using summary() function for showing summary of these variables. Also, WindDirection will be the name of this code.

```
 WindGustDir  WindGustSpeed    WindSpeed3pm    WindSpeed9am
NW      : 73   Min.    :13.00   Min.    : 0.00   Min.    : 0.000
NNW     : 44   1st Qu.:31.00   1st Qu.:11.00   1st Qu.: 6.000
E       : 37   Median :39.00   Median :17.00   Median : 7.000
WNW     : 35   Mean   :39.84   Mean   :17.99   Mean   : 9.652
ENE     : 30   3rd Qu.:46.00   3rd Qu.:24.00   3rd Qu.:13.000
(Other):144   Max.   :98.00   Max.   :52.00   Max.   :41.000
NA's    :  3   NA's   :2                       NA's   :7
```

Figure 19 output7

Figure above is the summary of the output for WindDirection. The variable that used for showing this output is WindGustDir, WindGustSpeed, WindSpeed3pm and WindSpeed9am. Also, the summary of these 4 variables has stated out for user to see.

1.3.8 Data Manipulation 8

```
weather %>% filter(RainToday == "Yes") %>% select(MinTemp, MaxTemp) %>% head(20)
```

Figure 20 manipulation 8

Figure above is the code for finding the highest temperature and lowest temperature when today is raining. The function used is Filter() and select() and head(). Head() function is used because the output have 66, and it is too long to show.

	MinTemp	MaxTemp
1	14.0	26.9
2	13.7	23.4
3	13.3	15.5
4	7.6	16.1
5	8.4	22.8
6	11.7	30.0
7	12.8	18.5
8	15.1	20.4
9	11.6	26.3
10	13.3	26.5
11	16.5	28.2
12	15.5	21.1
13	10.8	21.7
14	12.6	23.1
15	14.8	29.5
16	19.9	22.0
17	9.2	20.4
18	11.3	21.7
19	15.1	28.3
20	17.0	33.8

Figure 21 output 8

Figure above is output of minimum temperature and maximum temperature when today is raining. The output data have 66 but it just shows 20 data because of using head() function. Also, when today is raining, the MinTemp will not exceed 20 degree Celsius, and for MaxTemp, the degree Celsius will not exceed 34.

1.3.9 Data Manipulation 9

```
Humidity = weather %>% select(Humidity9am, Humidity3pm)
summary(Humidity)
```

Figure 22 manipulation 9

Figure above is use for selecting the humidity9am and 3pm. And show the summary of the variable.

```
  Humidity9am     Humidity3pm   
 Min.   :36.00   Min.   :13.00  
 1st Qu.:64.00   1st Qu.:32.25  
 Median :72.00   Median :43.00  
 Mean   :72.04   Mean   :44.52  
 3rd Qu.:81.00   3rd Qu.:55.00  
 Max.   :99.00   Max.   :96.00  
```

Figure 23 output 9

Figure above is the summary of the humidity9am and 3pm. It shows the minimal, 1st quarter, median, mean, 3rd quarter and maximum value in humidity9am and 3pm.

1.3.10 Data Manipulation 10

```
AverageHumidity =weather %>% mutate(AverageHumidity = (Humidity9am+Humidity3pm)/2) %>% select(AverageHumidity)
head(AverageHumidity, 20)
```

Figure 24 manipulation 10

Figure above is the code for finding the average of humidity by combining the humidity9am and 3pm. And show the first 20 output by using head() function.

	AverageHumidity
1	48.5
2	58.0
3	75.5
4	59.0
5	58.5
6	63.5
7	55.0
8	61.0
9	59.0
10	57.0
11	54.0
12	44.5
13	45.5
14	43.5
15	30.5
16	46.0
17	47.5
18	42.5
19	42.5
20	44.0

Figure 25 output 10

Figure above is the output for the code. It show the average of the sum of humidity9am and humidity3pm.

1.3.11 Data Manipulation 11

```
AveragePressure =weather %>% mutate(AveragePressure = (Pressure9am+Pressure3pm)/2) %>% select(AveragePressure, Pressure9am, Pressure3pm)
head(AveragePressure,20)
```

Figure 26 manipulation 11

Figure above is the code for name and find the average pressure by using mutate() function. And select the variable that I want to show which is AveragePressure, Pressure9am and Pressure3pm. After that show the first 20 output.

	AveragePressure	Pressure9am	Pressure3pm
1	1017.35	1019.7	1015.0
2	1010.40	1012.4	1008.4
3	1008.35	1009.5	1007.2
4	1006.25	1005.5	1007.0
5	1018.40	1018.3	1018.5
6	1022.75	1023.8	1021.7
7	1023.40	1024.6	1022.2
8	1025.20	1026.2	1024.2
9	1024.40	1026.1	1022.7
10	1022.40	1024.1	1020.7
11	1022.75	1024.4	1021.1
12	1021.85	1023.8	1019.9
13	1019.55	1022.0	1017.1
14	1015.20	1017.3	1013.1
15	1015.95	1018.2	1013.7
16	1015.35	1017.9	1012.8
17	1012.10	1014.4	1009.8
18	1014.70	1016.4	1013.0
19	1015.20	1017.1	1013.3
20	1016.10	1018.5	1013.7

Figure 27 output 11

Figure above is the output of the AveragePressure and Pressure9am and 3pm.

1.3.12 Data Manipulation 12

```
weather %>%
  group_by(RainToday) %>%
  summarise(AVGmaxTemp =mean(MaxTemp), AVGminTemp = mean(MinTemp)) %>%
filter(RainToday == "Yes")
```

Figure 28 manipulation 12

Figure above is the code of find the average max and min temperature when today is raining.

	RainToday	AVGmaxTemp	AVGminTemp
	<fct>	<dbl>	<dbl>
1	Yes	19.6	10.2

Figure 29 output 12

Figure above is the average max and min temperature when today is raining. For the average max temperature is 19.6 degree Celsius, and for minimum temperature is 10.2 degree Celsius.

1.3.13 Data Manipulation 13

```
weather %>%  na.omit %>%filter(RainToday=="Yes") %>% mutate(AverageWindSpeed = (WindSpeed9am+WindSpeed3pm)/2)
%>%select(RainToday,AverageWindSpeed) %>% arrange(AverageWindSpeed)
```

Figure 30 manipulation 13

Figure above is the code for finding the average wind speed while today is raining by using mutate, select and arrange function. First, filter the RainToday and make the result is all “Yes”. After that find average wind speed by sum WindSpeed9am and WindSpeed3pm and divide by 2. Lastly, arrange it by ascending order.

	RainToday	AverageWindSpeed
1	Yes	5.5
2	Yes	6.0
3	Yes	6.0
4	Yes	6.5
5	Yes	6.5
6	Yes	6.5
7	Yes	7.5
8	Yes	7.5
9	Yes	8.5
10	Yes	8.5
11	Yes	9.0
12	Yes	9.5
13	Yes	10.0
14	Yes	10.0
15	Yes	10.5
16	Yes	11.0
17	Yes	11.0
18	Yes	11.0
19	Yes	12.0
20	Yes	13.0

Figure 31 output 13

21	Yes	13.0
22	Yes	13.0
23	Yes	13.0
24	Yes	13.0
25	Yes	13.5
26	Yes	14.0
27	Yes	14.0
28	Yes	14.0
29	Yes	14.5
30	Yes	14.5
31	Yes	14.5
32	Yes	15.0
33	Yes	15.5
34	Yes	16.5
35	Yes	18.0
36	Yes	18.5
37	Yes	18.5
38	Yes	18.5
39	Yes	19.5
40	Yes	19.5
41	Yes	19.5

Figure 32 output 13.1

42	Yes	19.5
43	Yes	20.5
44	Yes	23.0
45	Yes	23.5
46	Yes	24.0
47	Yes	24.0
48	Yes	24.0
49	Yes	24.0
50	Yes	25.5
51	Yes	26.0
52	Yes	27.0
53	Yes	28.5
54	Yes	28.5
55	Yes	29.5
56	Yes	29.5
57	Yes	32.5
58	Yes	33.0
59	Yes	34.5
60	Yes	35.5
61	Yes	41.5

Figure 33 output 13.2

Figures above is the output of the code. As the graph shown, the lowest average wind speed from wind speed 9am and 3pm in raining day is 5.5 km/hr. And the highest average wind speed in 9am and 3pm is 41.5km/hr.

1.3.14 Data Manipulation 14

```
weather %>% na.omit %>%filter(Evaporation>6) %>% mutate(AverageCloud = (Cloud9am+Cloud3pm)/2)%>%
  select(Evaporation, AverageCloud) %>% arrange(Evaporation) %>% head(40)
```

Figure 34 manipulation 14

Figure above is the code for finding average of cloud9am and 3pm when the evaporation rate is over 6 and arrange evaporation as ascending order.

	Evaporation	AverageCloud
1	6.2	2.5
2	6.2	5.0
3	6.2	6.5
4	6.2	1.0
5	6.2	1.0
6	6.2	0.5
7	6.2	5.5
8	6.4	8.0
9	6.4	8.0
10	6.4	1.0
11	6.4	4.5
12	6.4	1.5
13	6.4	3.5
14	6.4	2.5
15	6.4	3.5
16	6.4	4.0
17	6.6	7.0
18	6.6	7.0
19	6.6	7.0
20	6.6	6.5
21	6.6	1.0
22	6.6	1.0
23	6.6	4.5
24	6.6	1.0
25	6.6	2.5
26	6.6	6.5
27	6.6	1.0
28	6.6	2.0
29	6.8	3.0
30	6.8	4.5
31	6.8	7.0
32	6.8	7.0
33	6.8	4.5
34	6.8	0.5
35	6.8	4.5
36	7.0	7.0
37	7.2	4.5
38	7.2	1.5
39	7.2	0.5
40	7.2	6.5
41	7.2	3.0
42	7.2	7.0
43	7.2	1.0
44	7.2	4.5
45	7.2	0.5
46	7.4	2.0
47	7.4	4.5
48	7.4	3.0
49	7.4	4.0
50	7.4	1.5

Figure 35 output 14

51	7.6	0.5
52	7.6	7.0
53	7.6	1.0
54	7.6	2.0
55	7.8	4.0
56	7.8	7.0
57	7.8	1.0
58	7.8	3.5
59	7.8	0.5
60	8.0	0.5
61	8.0	8.0
62	8.2	5.5
63	8.2	7.5
64	8.4	1.5
65	8.4	4.0
66	8.4	2.5
67	8.6	4.0
68	8.8	0.5
69	8.8	3.0
70	8.8	3.0
71	8.8	6.5
72	8.8	4.0
73	9.0	1.0
74	9.0	2.5
75	9.0	2.5
76	9.0	6.0
77	9.2	8.0
78	9.2	3.5
79	9.4	0.5
80	9.4	0.5
81	9.4	3.5
82	9.4	1.0
83	9.4	6.0
84	9.6	1.0
85	9.6	8.0
86	9.6	2.0
87	9.6	4.5
88	10.0	4.0
89	10.0	0.5
90	10.0	1.0
91	10.2	7.5
92	10.4	7.0
93	10.4	6.5
94	10.4	1.0
95	10.4	0.5
96	11.4	0.5
97	11.6	0.5
98	12.4	7.0
99	12.6	7.5
100	13.8	4.5

Figure 36 output 14.1

Figure above is the output for the code. It shows the evaporation and the average of the cloud. When the sky is fully obscured by clouds which is 8oktas, the evaporation(mm) in the 24hours to 9am is 9.2, 6.4, 9.6, and 8.0.

1.3.15 Data Manipulation 15

```
weather %>% filter(RainToday =="Yes", WindGustDir =="E", WindGustSpeed >40) %>% select(RainToday, WindGustDir, WindGustSpeed)
```

Figure 37 manipulation 15

Figure above is the code for find rainy day when wind direction is east when wind gust speed is over 40(km/h).

	RainToday	WindGustDir	WindGustSpeed
1	Yes	E	41
2	Yes	E	46

Figure 38 output 15

Figure above is output for rainy day when wind direction is east when wind gust speed is over 40(km/h). The figure shows that there is only 2 output which is fulfil the criteria.

1.3.16 Data Manipulation 16

```
weather %>% filter(RainTomorrow == "Yes") %>% select(MinTemp)
```

Figure 39 manipulation 16

	MinTemp
1	8.0
2	14.0
3	13.7
4	13.3
5	8.8
6	13.8
7	16.4
8	13.6
9	15.1
10	16.6
11	17.2
12	10.1
13	15.5
14	12.8
15	12.6
16	14.8
17	19.9
18	12.4
19	14.3
20	20.9
21	17.9
22	15.1
23	15.3
24	18.0
25	17.6
26	17.1
27	18.2
28	14.5
29	8.9
30	14.5
31	12.0
32	16.3
33	14.8
34	10.8
35	15.5
36	13.1
37	5.3
38	6.1
39	7.1
40	7.9

Figure 40 output 16.1

41	-0.4
42	-0.3
43	4.7
44	9.8
45	7.4
46	8.6
47	4.0
48	0.5
49	3.0
50	1.8
51	1.3
52	-1.1
53	-2.3
54	3.0
55	-1.9
56	-0.6
57	0.5
58	6.1
59	9.0
60	13.1
61	8.7
62	3.2
63	16.8
64	14.4
65	11.2
66	6.7

Figure 41 output16

Figure above show the MinTemp when tomorrow is going to rain. From the figure we can know when tomorrow will be going to rain, the temperature will not be very high.

1.3.17 Data Manipulation 17

```
weather %>% mutate(AverageCloud = (Cloud9am+Cloud3pm)/2) %>%
filter(Sunshine > 6) %>% select(AverageCloud, Sunshine) %>% arrange
(AverageCloud)
```

Figure 42 manipulation 17

Figure above is the code for average cloud when Sunshine is more than 6 hours and show it in ascending order.

	AverageCloud	Sunshine
1	0.0	10.4
2	0.0	10.9
3	0.0	11.0
4	0.5	13.0
5	0.5	13.1
6	0.5	10.4
7	0.5	13.6
8	0.5	10.5
9	0.5	13.3
10	0.5	13.2
11	0.5	12.4
12	0.5	11.5
13	0.5	11.3
14	0.5	11.2
15	0.5	10.7
16	0.5	10.3
17	0.5	9.6
18	0.5	10.6
19	0.5	10.2
20	0.5	9.5
21	0.5	9.2
22	0.5	8.9
23	0.5	10.4
24	0.5	10.9
25	0.5	11.0
26	0.5	11.5
27	0.5	11.5
28	0.5	12.7

Figure 43 output 17

29	1.0	13.0
30	1.0	13.6
31	1.0	13.0
32	1.0	10.5
33	1.0	12.6
34	1.0	12.8
35	1.0	11.6
36	1.0	11.8
37	1.0	12.7
38	1.0	12.6
39	1.0	12.6
40	1.0	12.4
41	1.0	12.1
42	1.0	11.6
43	1.0	10.1
44	1.0	11.0
45	1.0	11.2
46	1.0	9.5
47	1.0	9.3
48	1.0	10.6
49	1.0	9.4
50	1.0	9.4
51	1.0	9.4
52	1.0	9.3
53	1.0	9.3
54	1.0	9.2
55	1.0	8.5
56	1.0	8.6

Figure 44 17.1

57	1.0	9.0
58	1.0	9.0
59	1.0	8.5
60	1.0	8.9
61	1.0	8.9
62	1.0	9.2
63	1.0	8.7
64	1.0	9.3
65	1.0	9.6
66	1.0	9.9
67	1.0	9.8
68	1.0	9.4
69	1.0	10.0
70	1.0	9.5
71	1.0	10.8
72	1.0	9.4
73	1.0	9.9
74	1.0	10.3
75	1.0	11.3
76	1.0	11.1
77	1.0	11.1
78	1.0	10.8
79	1.0	11.8
80	1.0	12.2
81	1.0	12.1
82	1.0	12.6
83	1.5	11.9
84	1.5	12.5
85	1.5	11.1
86	1.5	12.7
87	1.5	11.6
88	1.5	12.1

Figure 45 17.2

88	1.5	12.1
89	1.5	8.5
90	1.5	9.6
91	1.5	10.8
92	1.5	10.1
93	1.5	9.2
94	1.5	9.4
95	1.5	8.7
96	1.5	8.5
97	1.5	8.2
98	1.5	7.0
99	1.5	8.1
100	1.5	9.7
101	1.5	11.4
102	1.5	11.0
103	1.5	12.2
104	1.5	10.8
105	2.0	13.0
106	2.0	12.0
107	2.0	11.0
108	2.0	9.7
109	2.0	9.2
110	2.0	9.0
111	2.0	8.1
112	2.0	8.6
113	2.0	8.4
114	2.0	10.2
115	2.0	11.0
116	2.0	11.2
117	2.0	11.7
118	2.0	12.1
119	2.5	12.4

Figure 46 output 17.3

244	6.5	8.8
245	6.5	8.7
246	6.5	12.8
247	6.5	12.5
248	6.5	8.4
249	6.5	9.2
250	6.5	6.2
251	6.5	8.0
252	6.5	8.9
253	6.5	6.1
254	6.5	6.8
255	6.5	7.0
256	6.5	7.8
257	6.5	7.1
258	7.0	6.3
259	7.0	10.6
260	7.0	11.1
261	7.0	10.0
262	7.0	7.1
263	7.0	8.2
264	7.0	8.0
265	7.0	6.5
266	7.0	6.8
267	7.0	8.6
268	7.5	6.3
269	7.5	8.1
270	7.5	6.2

Figure 47 output 17.4

Figures above is the output for average cloud when Sunshine is more than 6 hours. The average cloud is showing in ascending order. When the average cloud is 0 oktas, the sunshine is 10.4 hours in the day. Also, when average cloud is 7.5 oktas, the sunshine is 6.2 hours in the day.

1.3.18 Data Manipulation 18

```
PredictRisk1 = weather %>% select(RainToday, RainTomorrow, RISK_MM) %>%
filter(RainToday =="Yes", RainTomorrow =="Yes")
PredictRisk2 = weather %>% select(RainToday, RainTomorrow, RISK_MM)%
>%filter(RainToday =="No", RainTomorrow =="No")
```

Figure 48 manipulation 18

Figure above is the code to show prediction of risk happening tomorrow by selecting the RISK_MM, RainToday and RainTomorrow. To compare, I've filtered the data with no rain for today and tomorrow and got rain for today and tomorrow.

```
> head(PredictRisk1)
  RainToday RainTomorrow RISK_MM
1       Yes          Yes    3.6
2       Yes          Yes   39.8
3       Yes          Yes    2.8
4       Yes          Yes    4.2
5       Yes          Yes    1.4
6       Yes          Yes    6.4
> head(PredictRisk2)
  RainToday RainTomorrow RISK_MM
1        No           No    0.2
2        No           No    0.0
3        No           No    0.0
4        No           No    0.2
5        No           No    0.0
6        No           No    0.0
```

Figure 49 output 18

Figure above is the risk predict. As the figure shown, when it going to be rain, the RISK_MM will be more than 1.0. Besides, when it not going to rain, the RISK_MM will not be going to be more than 1.0.

1.3.19 Data Manipulation 19

```
weather %>% group_by(RainToday) %>% filter(RainToday =="Yes") %>%
summarise(AVGrainfall = mean(Rainfall), AVGRISKMM= mean(RISK_MM)) %
>%select(RainToday, AVGrainfall, AVGRISKMM )
```

Figure 50 manipulation 19

Figure above is the code for showing average rainfall and average RISK_MM when today is raining. It is to see is RISK_MM have the accurate data a not.

```
  RainToday AVGrainfall AVGRISKMM
  <fct>           <dbl>     <dbl>
1 Yes              7.66      2.79
>
```

Figure 51 output19

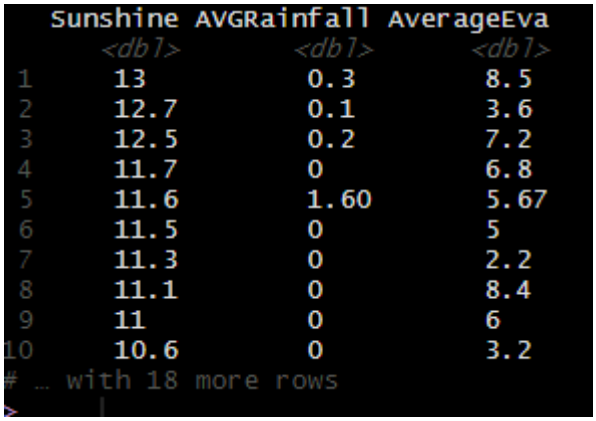
Figure above is the output of raintoday, AVGrainfall and AVGRISKMM. As the figure show, we can see the average rainfall is more than average RISK_MM.

1.3.20 Data Manipulation 20

```
weather%>% group_by(Sunshine) %>% select(Sunshine, WindGustDir, Rainfall,
Evaporation) %>% filter(WindGustDir =="E") %>% summarise(AVGRainfall =
mean(Rainfall), AverageEva = mean(Evaporation)) %>%  arrange(desc(Sunshine
))
```

Figure 52 manipulation 20

Figure above is the code for summarize the Average of Rainfall, average of evaporation for each wind direction that face "E". And then order the output by the Sunshine in descending order. It is use for check what is the average of rainfall and evaporation when the highest bright sunshine in a day.



	Sunshine	AVGRainfall	AverageEva
	<dbl>	<dbl>	<dbl>
1	13	0.3	8.5
2	12.7	0.1	3.6
3	12.5	0.2	7.2
4	11.7	0	6.8
5	11.6	1.60	5.67
6	11.5	0	5
7	11.3	0	2.2
8	11.1	0	8.4
9	11	0	6
10	10.6	0	3.2

… with 18 more rows

Figure 53 output 20

Figure above is the output of the code. It has clearly stated that when there's sunshine, the average rainfall will not exceed 2. Also, the average of evaporation will also be high when having long hours of sunshine when the wind facing East.

1.3.21 Data Manipulation 21

```
weather %>% group_by(Sunshine) %>% select(Sunshine, WindGustDir, MinTemp,
MaxTemp) %>% filter(WindGustDir =="E") %>% summarise(AVGMinTemp = mean
(MinTemp, na.rm = TRUE), AVGMaxTemp = mean(MaxTemp, na.rm = TRUE)) %>%
arrange(Sunshine)
```

Figure 54 manipulation 21

Figure above is the code for summarise the Average of MinTemp, Average of MaxTemp for the wind facing "E". The average temperature on 3pm and 9am and order the output by Sunshine in ascending order.

	Sunshine	AVGMinTemp	AVGMaxTemp
	<dbl>	<dbl>	<dbl>
1	0	16.4	19.4
2	1.6	11.9	19.0
3	3	4.3	14.5
4	4.6	8.3	17
5	5.6	11.3	21.7
6	5.9	-0.2	16.2
7	6	16	22.8
8	7.3	3.8	17.4
9	7.5	10.8	29.2
10	7.7	8.4	22.8

… with 18 more rows

Figure 55 output 21

Figure above is the output for the code. From the figure we can know that the sunshine does not affect the temperature because when the hours of sunshine is 0, the minimum and maximum temperature are similar.

1.3.22 Data Manipulation 22

```
weather  %>% filter(WindDir9am =="E", WindDir3pm == "E") %>% summarise
(MaxWindSpeed = max(WindGustSpeed))
```

Figure 56 manipulation 22

Figure above is the code for find the highest wind speed of the if the wind at 9am and 3pm is facing to East.

	MaxWindSpeed
1	41

>

Figure 57 output 22

Figure above is the output of maximum wind speed when the wind direction at 9am and 3 pm is facing to east.

1.3.23 Data Manipulation 23

```
AVGtemperatureNoRain = weather %>%  filter(RainToday =="No") %>% mutate
(AVGtempertature = (MaxTemp+MinTemp)/2) %>% select(RainToday, AVGtempertat
ure)
head(AVGtemperatureNoRain, 10)
```

Figure 58 manipulation 23

Figure above is the code for finding the average of minimum and maximum temperature of the day when there is not raining.

	RainToday	AVGtempertature
1	No	16.15
2	No	11.55
3	No	12.15
4	No	12.65
5	No	14.15
6	No	17.15
7	No	17.90
8	No	19.00
9	No	21.50
10	No	20.65

Figure 59 output 23

Figure above is the output for the average of minimum and maximum temperature of the day when there is not raining. From the figure we can know that the average temperature when not raining doesn't have big difference.

1.3.24 Data Manipulation 24

```
weather %>% filter(RainToday =="No", RainTomorrow =="Yes") %>% summarise
(MinT = min(MinTemp,na.rm = TRUE), Min9 = min(Temp9am,na.rm = TRUE), Min3
= min(Temp3pm,na.rm = TRUE))
```

Figure 60 manipulation 24

Figure above is the code for find the minimum of MinTemp, temperature at 9am "and 3 pm when tomorrow is raining and today is not raining.

```
  MinT Min9 Min3
1 -2.3    1  5.7
>
```

Figure 61 output 24

Figure above is the output of the minimum of MinTemp, temperature at 9am "and 3 pm when tomorrow is raining and today is not raining. From the figure, we can know when tomorrow is raining, the temperature usually will be lower.

1.3.25 Data Manipulation 25

```
weather %>% filter(RainToday =="No", RainTomorrow =="Yes") %>% summarise
(MinC = min(WindGustSpeed,na.rm = TRUE), Min9 = min(WindSpeed9am,na.rm =
TRUE), Min3 = min(WindSpeed3pm,na.rm = TRUE))
```

Figure 62 manipulation 25

Figure above is the code for find the minimum of WindGustSpeed, WindSpeed9am at 9am and 3 pm when tomorrow is raining and today is not raining.

```
  MinC Min9 Min3
1   17    0    6
```

Figure 63 output 25

Figure above is the output for the code. It shows that the minimum of WindGustSpeed, WindSpeed9am at 9am and 3 pm when tomorrow is raining and today is not raining. From the figure, when tomorrow is raining, the minimum wind speed is 17km/h, but not in 9am, is after 3pm.

1.4 Data Visualization and Exploration

1.4.1 Data Visualization and Exploration 1

```
plot(x=weather$Evaporation)
```

Figure 64 Data Visualization and Exploration 1

Figure above is the code for plot the graph of evaporation.

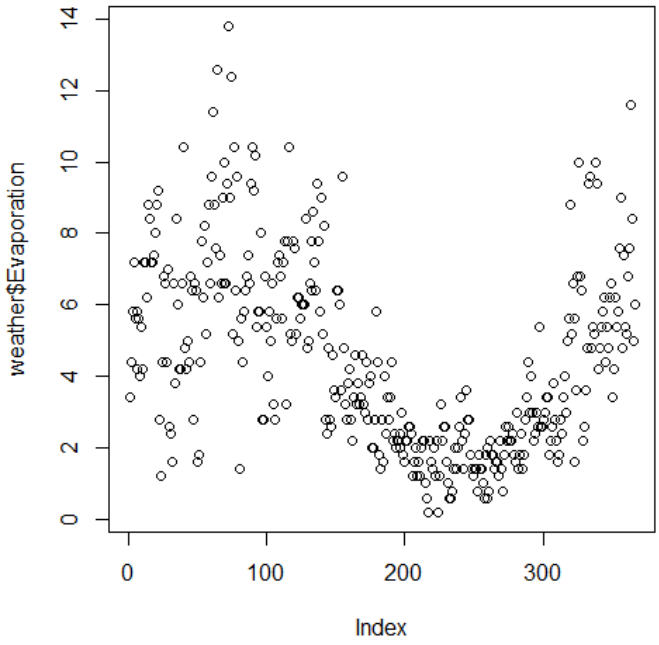


Figure 65 output Data Visualization and Exploration 1

Figure above is the output for evaporation. As the graph show, we can know the evaporation have highest rate at 13.8, lowest rate at 0.2 .

1.4.2 Data Visualization and Exploration 2

```
ggplot(weather, aes(x=WindGustSpeed)) + geom_histogram(col="white", fill
="blue")
```

Figure 66 Data Visualization and Exploration 2

Figure above is the code for plot the histogram of WindGustSpeed.

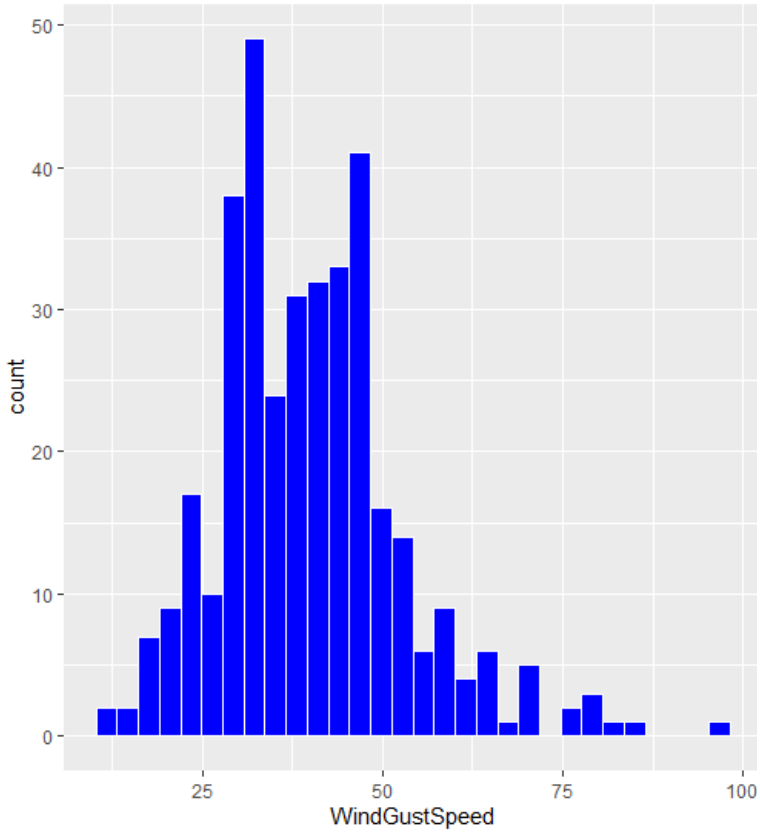


Figure 67 output Data Visualization and Exploration 2

Figure above is the graph of histogram of WindGustSpeed. From the graph above, we can know the highest count of WindGustSpeed is 48 and lowest count is 0.

1.4.3 Data Visualization and Exploration 3

```
ggplot(data = weather) + geom_bar(mapping = aes(x = WindGustDir))
```

Figure 68 Data Visualization and Exploration 3

Figure above is the code for plot bar chart for direction of wind.

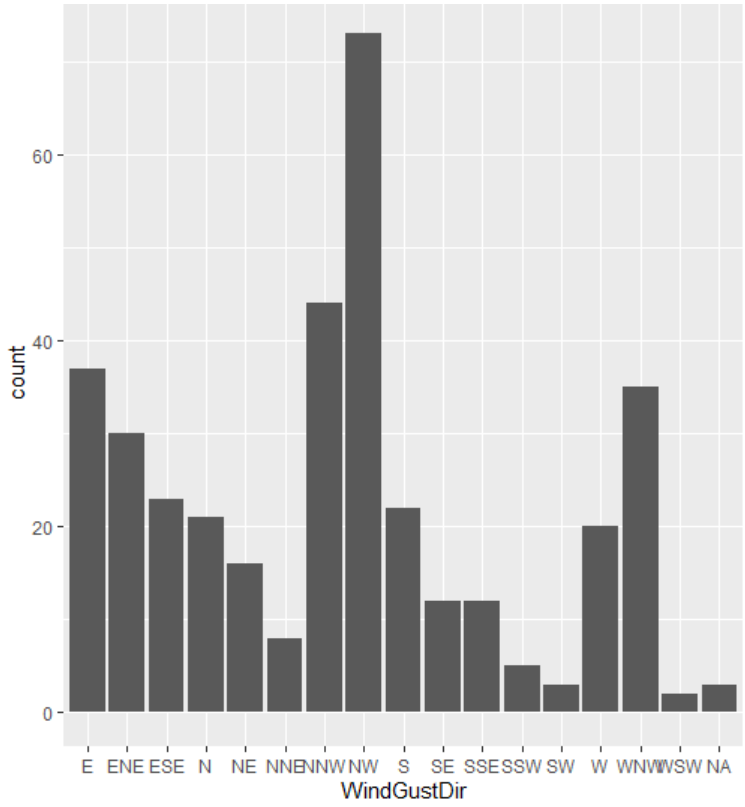


Figure 69 output Data Visualization and Exploration 3

Figure above is the bar chart of direction of wind. From the graph, we know the NW have the highest count and WSW have the lowest count.

1.4.4 Data Visualization and Exploration 4

```
ggplot(na.omit(weather),aes(y=Humidity9am, x=RainToday)) + geom_boxplot
(outlier.shape=NA)
```

Figure 70 Data Visualization and Exploration 4

Figure above is the code for plot boxplot of humidity9am and RainToday. na.omit is used for removing NA value.

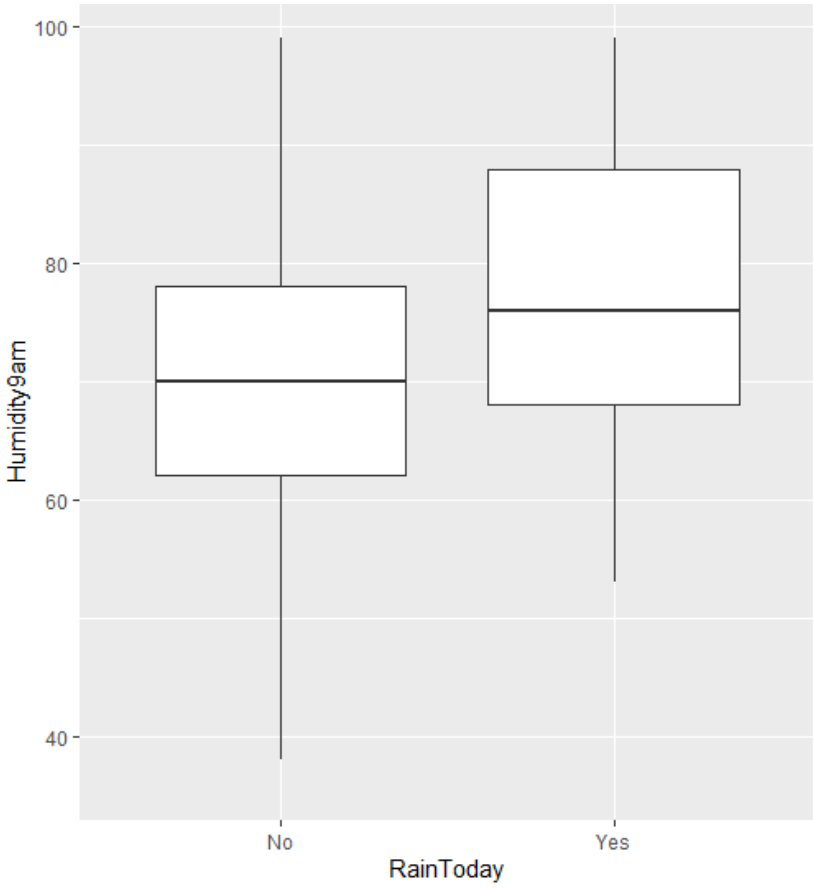


Figure 71 output Data Visualization and Exploration 4

Figure above is the boxplot for humidity9am and Rain today. From the graph, we can know when the RainToday is yes, the humidity at 9am is higher than today doesn't rain.

1.4.5 Data Visualization and Exploration 5

```
ggplot(na.omit(weather),aes(y=Evaporation, x=Pressure9am)) + geom_line()
```

Figure 72 Data Visualization and Exploration 5

Figure above is the code for plot line graph. The x-axis is pressure9am, y-axis is evaporation.

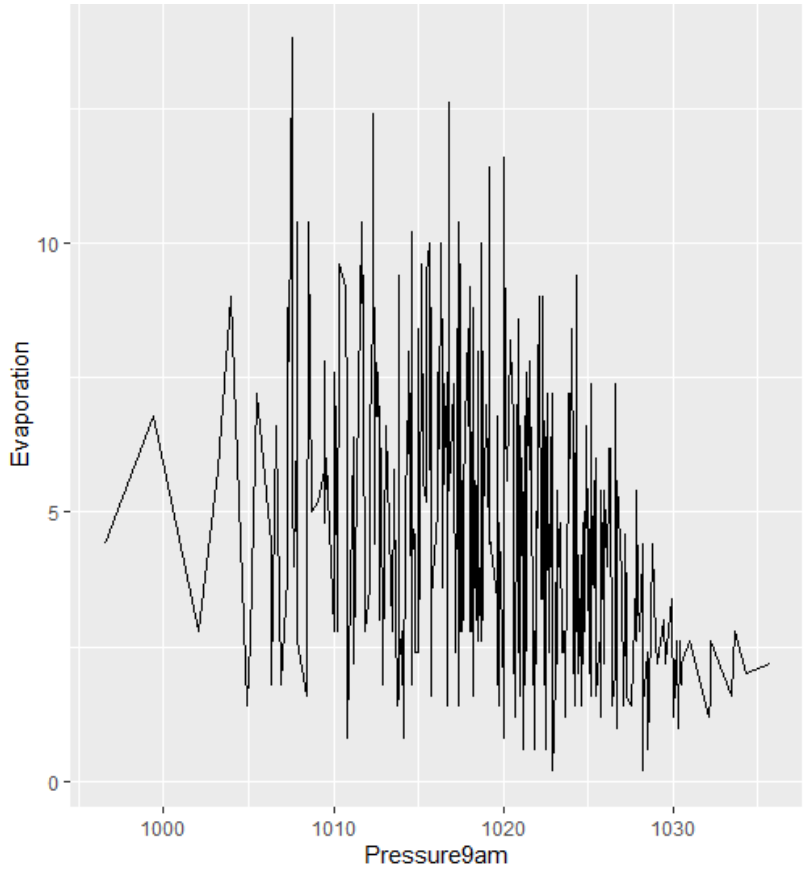


Figure 73 output Data Visualization and Exploration 5

Figure above is the line graph of evaporation and the pressure at 9am. As the graph show, the evaporation is high when the pressure is at 1008.

1.4.6 Data Visualization and Exploration 6

```
ggplot(data = weather) + geom_point(mapping = aes(x = MaxTemp, y =
Evaporation))+ labs(title = "Scatterplot between Evaporation and MaxTemp")
```

Figure 74 1.5.1 Data Visualization and Exploration 6

Figure above is the code for scatterplot for evaporation and max temperature.

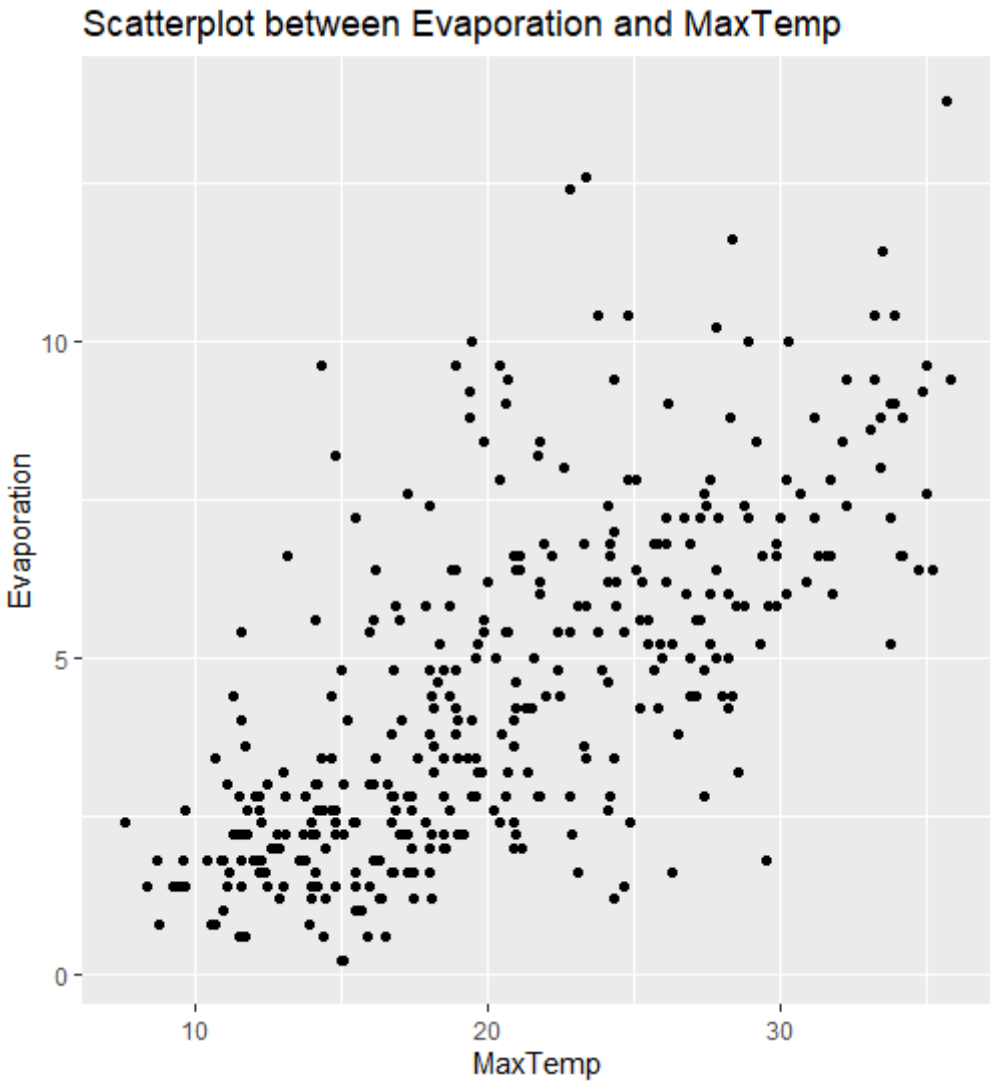


Figure 75 output Data Visualization and Exploration 6

Figure above is the scatterplot for finding the relationship between evaporation and max temperature. From this graph, we can see the max temperature can affect the evaporation. When the temperature is higher the evaporation will also be high.

1.4.7 Data Visualization and Exploration 7

```
ggplot(na.omit(weather)) + geom_point(mapping = aes(x = MaxTemp, y = Sunshine))
```

Figure 76 1.5.1 Data Visualization and Exploration 7

Figure above is the point plot for sunshine and max temperature.

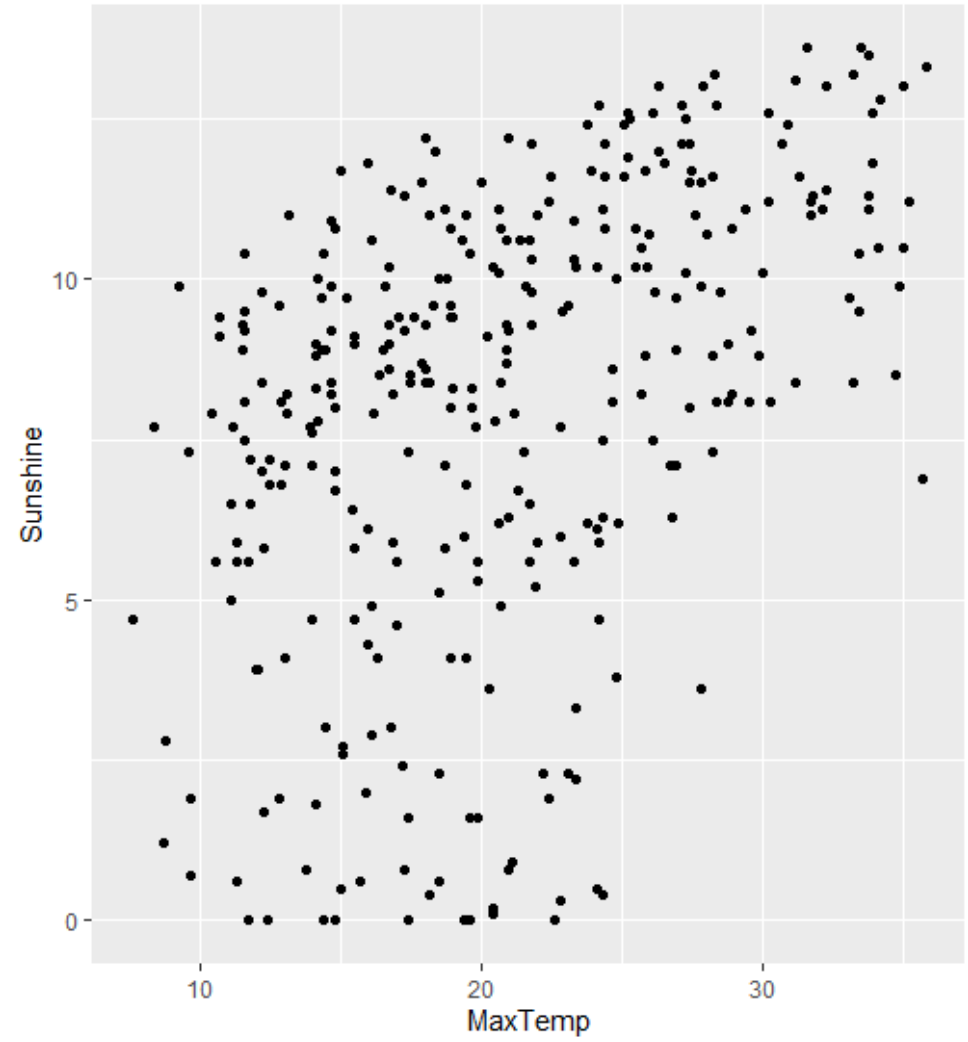


Figure 77output Data Visualization and Exploration 7

Figure above is the point graph of sunshine and max temperature. From the graph, we know that when sunshine stays longer, the temperature will be higher.

1.4.8 Data Visualization and Exploration 8

```
ggplot(data = ContinueRain) + geom_bar(mapping=aes(x = WindGustDir, fill
=WindGustDir)) + labs(xlab = "Wind Direction")
```

Figure 78 1.5.1 Data Visualization and Exploration 8

Figure above is the code for plot a bar chart for ContinueRain which is today and tomorrow will rain.

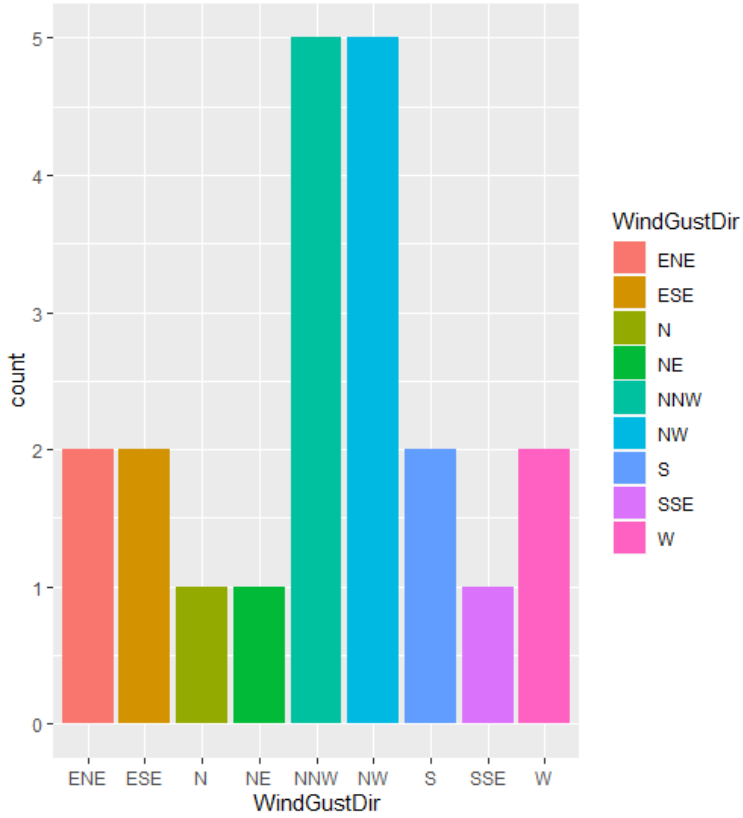


Figure 79output Data Visualization and Exploration 8

Figure above show the bar chart of wind direction when today and tomorrow is rain.

1.4.9 Data Visualization and Exploration 9

```
ggplot(data =NoRain)+  geom_bar(mapping=aes(x = WindGustDir, fill=WindGust
Dir)) + labs(xlab = "Wind Direction")
```

Figure 80 1.5.1 Data Visualization and Exploration 9

Figure above is the code for plot a bar chart of the wind direction.

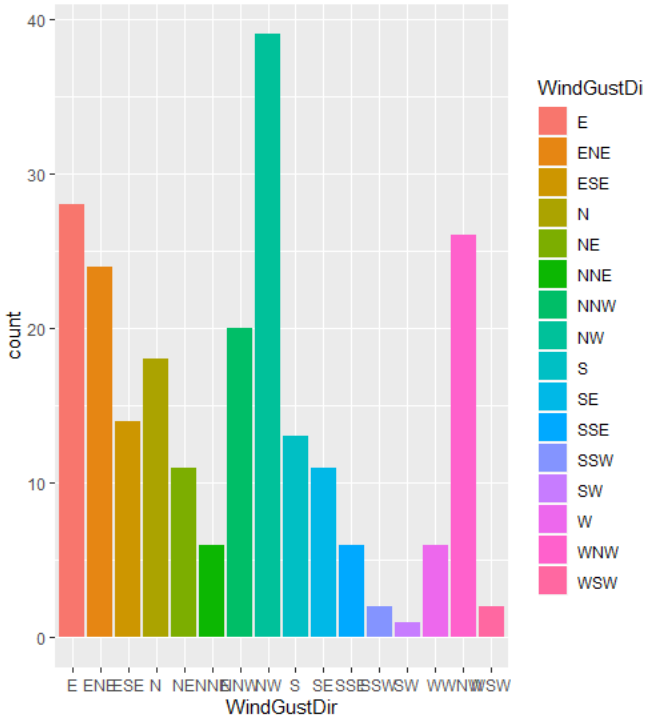


Figure 81output Data Visualization and Exploration 9

Figure above is the bar chart of wind direction. It clearly states the wind direction from the dataset.

1.4.10 Data Visualization and Exploration 10

```
#5
ggplot(na.omit(weather), mapping = aes(x = RainToday, y = Sunshine, )) +
geom_boxplot()
```

Figure 82 1.4.10 Data Visualization and Exploration 10

Figure above is the boxplot for RainToday and sunshine.

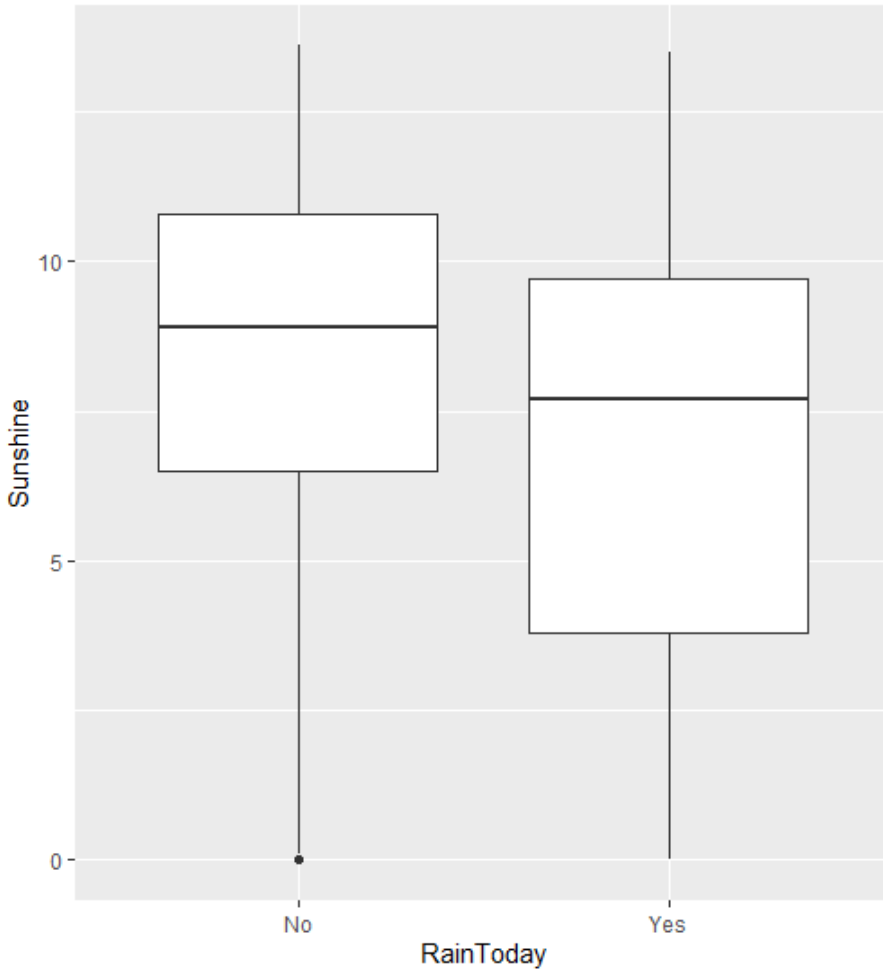


Figure 83 output Data Visualization and Exploration 10

Figure above is the graph for Sunshine and RainToday. When today doesn't rain, the sunshine will stay longer, and when today is raining, the sunshine will be shorter than when it is raining.

Conclusion

In conclusion, in this assignment, I have learnt many techniques for R programming. R programming is a very good program to analysis data. In this assignment, I have applied manipulation, exploration, and visualization by using the code like select(), mutate(), arrange(), filter(), group_by() and summarize() for data manipulation. And for exploration and visualization, I have used ggplot function for create the chart. In order to use the function like ggplot and select(). In order to use those function, I have installed library like ggplot2, crayon and dplyr. These libraries have very useful function for analysis the dataset.

Table of Figure

References

Naveen. (4 12, 2019). *intellipaat*. Retrieved from intelipaat: https://intellipaat.com/blog/tutorial/r-programming/data-manipulation-in-r/

rdocumentation. (12 02, 2020). *rdocumentation*. Retrieved from rdocumentation: https://www.rdocumentation.org/packages/rattle/versions/5.3.0/topics/weather

RDocumentation. (16 02, 2020). *rdocumentation*. Retrieved from rdocumentation: rdocumentation