ASSIGNMENT 2 STATISTICAL CLIMATOLOGY		
Student's Code	AIMS African Institute for Mathematical Sciences CAMEROON	Deadline
i7BR62FMv		**29.03.20, 20:00**
March 30, 2020		2019-2020
Lecturers: Roger Stern, James Musyoka and Danny Parsons		

1 Download Satellite data

In this part, it is a question of choosing a location in the world and downloading the corresponding climate data. For this, we are interested in Cameroon more precisely in the city of Garoua. Garoua is a city with a tropical climate. It would be interesting to study the distribution of precipitation in this area. To download our data we used the source CHIPRS and following the steps:

- We first go to Google map to get our coordinates (longitudes and latitudes).
- After enter it in CHIRPS platform and it look like this:

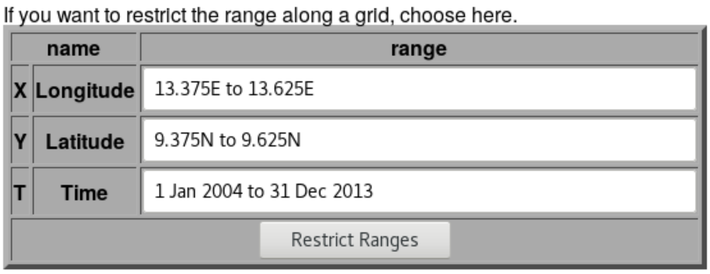

Figure 1: setting ranges from CHIRPS source

After Downloaded the data, this is how it appear in R-instat:

	X	Y	.T	requested_X	requested_Y	prcp	date
1	13.38	9.375	2453006	13.4	9.38	0.000e+00	2004-01-01
2	13.38	9.375	2453007	13.4	9.38	0.000e+00	2004-01-02
3	13.38	9.375	2453008	13.4	9.38	0.000e+00	2004-01-03
4	13.38	9.375	2453009	13.4	9.38	0.000e+00	2004-01-04
5	13.38	9.375	2453010	13.4	9.38	0.000e+00	2004-01-05
6	13.38	9.375	2453011	13.4	9.38	0.000e+00	2004-01-06
7	13.38	9.375	2453012	13.4	9.38	0.000e+00	2004-01-07
8	13.38	9.375	2453013	13.4	9.38	0.000e+00	2004-01-08
9	13.38	9.375	2453014	13.4	9.38	0.000e+00	2004-01-09
10	13.38	9.375	2453015	13.4	9.38	0.000e+00	2004-01-10
11	13.38	9.375	2453016	13.4	9.38	0.000e+00	2004-01-11
12	13.38	9.375	2453017	13.4	9.38	0.000e+00	2004-01-12
13	13.38	9.375	2453018	13.4	9.38	0.000e+00	2004-01-13
14	13.38	9.375	2453019	13.4	9.38	0.000e+00	2004-01-14
15	13.38	9.375	2453020	13.4	9.38	0.000e+00	2004-01-15
16	13.38	9.375	2453021	13.4	9.38	0.000e+00	2004-01-16
17	13.38	9.375	2453022	13.4	9.38	0.000e+00	2004-01-17
18	13.38	9.375	2453023	13.4	9.38	0.000e+00	2004-01-18
19	13.38	9.375	2453024	13.4	9.38	0.000e+00	2004-01-19

Figure 2: Satellite data for Garoua city from CHIRPS source

This data is about precipitation or rain in Garoua Cameroon from 2004 to 2013. We have choose this period of time because is was the one who contain much of data after trying several time to download the data.

Challenge in the process of downloading the satellite data

The way way to dowload satellite data is very different for the classic one.

- First you need to register on their website, that was the case for CM SAF.
- Too many steps: You need to get the coordinate in Google map
- One of the main challenge is that, you can not visualize your data before downloaded them.That why we were trying to download the same data many time until you have a data that can satisfied you.

Now we have our data, we can try to do some analysis we did with station data in the previous assignment.

1.1 PICSA graphs with satellite data

We can first look at the descriptive statistic by doing the summary table. To do it **Describe** > **One Variable** > **Summarise**, and we obtain the following table:

```
       X              Y               .T              requested_X    requested_Y
 Min.   :13.4   Min.   :9.38   Min.   :2453006   Min.   :13.4   Min.   :9.38
 1st Qu.:13.4   1st Qu.:9.38   1st Qu.:2453919   1st Qu.:13.4   1st Qu.:9.38
 Median :13.4   Median :9.38   Median :2454832   Median :13.4   Median :9.38
 Mean   :13.4   Mean   :9.38   Mean   :2454832   Mean   :13.4   Mean   :9.38
 3rd Qu.:13.4   3rd Qu.:9.38   3rd Qu.:2455745   3rd Qu.:13.4   3rd Qu.:9.38
 Max.   :13.4   Max.   :9.38   Max.   :2456658   Max.   :13.4   Max.   :9.38

      prcp             date                 year        month_abbr
 Min.   : 0.00   Min.   :2004-01-01   Min.   :2004   Jan    : 310
 1st Qu.: 0.00   1st Qu.:2006-07-02   1st Qu.:2006   Mar    : 310
 Median : 0.00   Median :2008-12-31   Median :2008   May    : 310
 Mean   : 2.57   Mean   :2008-12-31   Mean   :2008   Jul    : 310
 3rd Qu.: 2.28   3rd Qu.:2011-07-02   3rd Qu.:2011   Aug    : 310
 Max.   :57.44   Max.   :2013-12-31   Max.   :2013   Oct    : 310
                                                     (Other):1793
  day_in_month        doy
 Min.   : 1.0   Min.   :  1
 1st Qu.: 8.0   1st Qu.: 93
 Median :16.0   Median :184
 Mean   :15.7   Mean   :184
 3rd Qu.:23.0   3rd Qu.:275
 Max.   :31.0   Max.   :366
```

Figure 3: Summarise Table for Garoua data

The first thing we notice looking at this table is that there is no missing value. So our period of time was well choose. We can also notice that there is no much rain in Garoua. The maximum for precipitation was 57.44 and the minimum 0.

Now let look at the PICSA graphs to have a better vew on this data. Now lets go to **Climatic** > **PICSA** > **Rainfall Graph**. We will obtain the following graph.

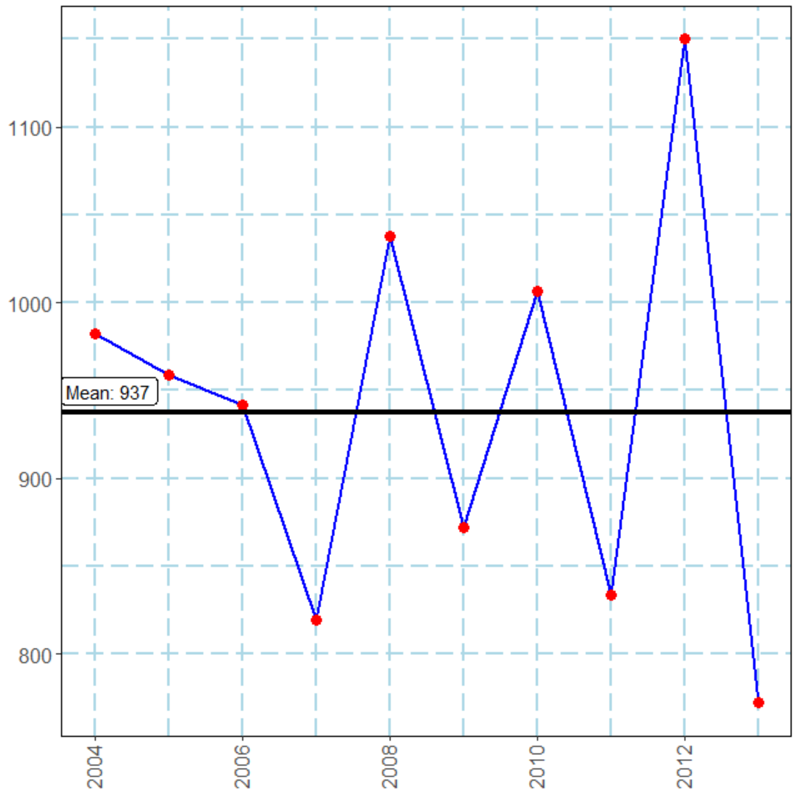

Figure 4: Sum of Precipitation for Garoua Satellite data

This graph above represent the sum of precipitation from 2004 to 2013. That's nice. We can see that even with satellite data, our graph looks like the one obtain with Station data.Looking at this graph, years 2012 and 2013 was the ones whose having much rain.

We can also look at the boxplot graph. Let's go to **Describe** > **Specific** > **Boxplot**. We obtain the following graph:

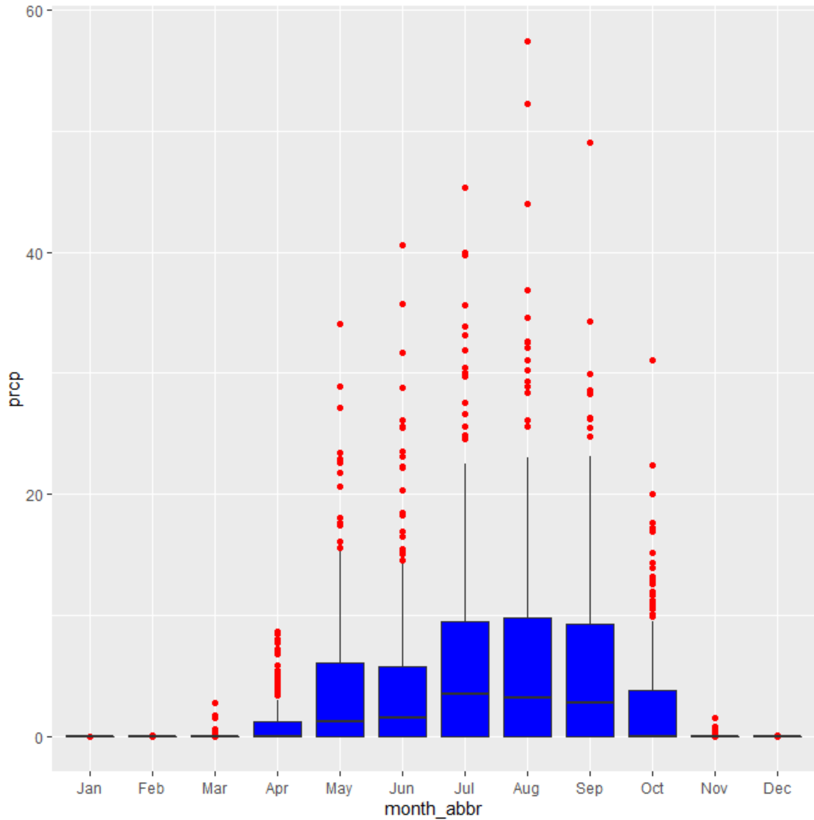

Figure 5: Boxplot of monthly rainfall totals for Garoua Satellite data

Now that we have our data and know that we can also do same analysis like with station data, lets compare have a station data for the same region and compare the two.

2 Comparison Station and Satellite Garoua Data in R-Instat

The main objective of this part is to see if it is possible to use the estimate satellite data for precipitation to complete the station data. Sp we first need to load our station data. To do this, we click on **File** > **Open From Library** and we choose our Garoua data.

	row.names	date	year	month	day	rain	tmin	tmax
1	18263	2004-01-01	2004	1	1	0.0	16.6	35.0
2	18264	2004-01-02	2004	1	2	0.0	18.6	36.0
3	18265	2004-01-03	2004	1	3	0.0	18.0	36.5
4	18266	2004-01-04	2004	1	4	0.0	19.0	36.2
5	18267	2004-01-05	2004	1	5	0.0	18.5	36.8
6	18268	2004-01-06	2004	1	6	0.0	20.6	38.3
7	18269	2004-01-07	2004	1	7	0.0	19.1	33.8
8	18270	2004-01-08	2004	1	8	0.0	18.5	30.2
9	18271	2004-01-09	2004	1	9	0.0	18.9	30.7
10	18272	2004-01-10	2004	1	10	0.0	19.8	29.4
11	18273	2004-01-11	2004	1	11	0.0	16.5	31.0
12	18274	2004-01-12	2004	1	12	0.0	16.8	31.8
13	18275	2004-01-13	2004	1	13	0.0	17.0	30.5
14	18276	2004-01-14	2004	1	14	0.0	17.0	29.7
15	18277	2004-01-15	2004	1	15	0.0	17.5	31.6
16	18278	2004-01-16	2004	1	16	0.0	18.5	33.0
17	18279	2004-01-17	2004	1	17	0.0	16.9	34.0
18	18280	2004-01-18	2004	1	18	0.0	17.6	34.0
19	18281	2004-01-19	2004	1	19	0.0	16.6	34.4

Figure 6: Vew of station Garoua data

After load the data, this is how it's look.
Now we will meerge the data station and satellite to have it as one table.

2.1 Merging the data

By merging the date, it would be easy for us to make different analysis to compare the two data. We are particular interesting on the rainfall, so for the satellite data we will just consider the precipitation column. To merge the data in R-instat, we click on **Prepare** > **Column: Reshape** > **Merge** and we enter the information in the dialogue box. By follow the different instruction we will obtain this:

	date	year	month	day	tmin	tmax	rain	prcp	year1
1	2004-01-01	2004	1	1	16.6	35.0	0.0	0.000e+00	2004
2	2004-01-02	2004	1	2	18.6	36.0	0.0	0.000e+00	2004
3	2004-01-03	2004	1	3	18.0	36.5	0.0	0.000e+00	2004
4	2004-01-04	2004	1	4	19.0	36.2	0.0	0.000e+00	2004
5	2004-01-05	2004	1	5	18.5	36.8	0.0	0.000e+00	2004
6	2004-01-06	2004	1	6	20.6	38.3	0.0	0.000e+00	2004
7	2004-01-07	2004	1	7	19.1	33.8	0.0	0.000e+00	2004
8	2004-01-08	2004	1	8	18.5	30.2	0.0	0.000e+00	2004
9	2004-01-09	2004	1	9	18.9	30.7	0.0	0.000e+00	2004
10	2004-01-10	2004	1	10	19.8	29.4	0.0	0.000e+00	2004
11	2004-01-11	2004	1	11	16.5	31.0	0.0	0.000e+00	2004
12	2004-01-12	2004	1	12	16.8	31.8	0.0	0.000e+00	2004
13	2004-01-13	2004	1	13	17.0	30.5	0.0	0.000e+00	2004
14	2004-01-14	2004	1	14	17.0	29.7	0.0	0.000e+00	2004
15	2004-01-15	2004	1	15	17.5	31.6	0.0	0.000e+00	2004
16	2004-01-16	2004	1	16	18.5	33.0	0.0	0.000e+00	2004
17	2004-01-17	2004	1	17	16.9	34.0	0.0	0.000e+00	2004
18	2004-01-18	2004	1	18	17.6	34.0	0.0	0.000e+00	2004
19	2004-01-19	2004	1	19	16.6	34.4	0.0	0.000e+00	2004

Figure 7: Merge Data for Garoua city

Just look at this table, we can notice that the 19 first rows are the same for **rain** and **prcp** columns. That's a good sign at first. But let's have the summarize table for a better opinion on this data.

2.2 Summaries Table

By doing the summarise table, we will have a view of all our descriptives statistic. To do this , let's go to **Describe** > **One Variable** > **Summarise** and we obtain this:

```
      date                 year          month            day     
 Min.   :2004-01-01   Min.   :2004   Min.   : 1.00   Min.   :  1  
 1st Qu.:2006-07-02   1st Qu.:2006   1st Qu.: 4.00   1st Qu.: 92  
 Median :2008-12-31   Median :2008   Median : 7.00   Median :183  
 Mean   :2008-12-31   Mean   :2008   Mean   : 6.52   Mean   :183  
 3rd Qu.:2011-07-02   3rd Qu.:2011   3rd Qu.:10.00   3rd Qu.:274  
 Max.   :2013-12-31   Max.   :2013   Max.   :12.00   Max.   :366  

      tmin            tmax            rain              prcp            year1     
 Min.   :11.7   Min.   :24.0   Min.   :  0.00   Min.   : 0.00   Min.   :2004  
 1st Qu.:20.6   1st Qu.:32.5   1st Qu.:  0.00   1st Qu.: 0.00   1st Qu.:2006  
 Median :23.0   Median :35.0   Median :  0.00   Median : 0.00   Median :2008  
 Mean   :22.7   Mean   :35.3   Mean   :  2.83   Mean   : 2.57   Mean   :2008  
 3rd Qu.:24.7   3rd Qu.:38.0   3rd Qu.:  0.00   3rd Qu.: 2.28   3rd Qu.:2011  
 Max.   :73.9   Max.   :54.0   Max.   :114.70   Max.   :57.44   Max.   :2013  
 NA's   :274    NA's   :2220   NA's   :3                                      
   month_abbr   day_in_month       doy     
 Jan    : 310   Min.   : 1.0   Min.   :  1  
 Mar    : 310   1st Qu.: 8.0   1st Qu.: 93  
 May    : 310   Median :16.0   Median :184  
 Jul    : 310   Mean   :15.7   Mean   :184  
 Aug    : 310   3rd Qu.:23.0   3rd Qu.:275  
 Oct    : 310   Max.   :31.0   Max.   :366  
 (Other):1793                               
```

Figure 8: Merge Data for Garoua city

According to this result, we can see that summary statistic for rain is different from summary statistic for prcp. Fisrt we have 3 missing values for rain column when for prcp there is no missing values. The **mean** for **rain** is **2.83** with **114.7** as a **maximum**; for **prcp**, we have **2.57** on average with **57.44** as a **maximum**. That is a great difference.

Now that we know our data are different, lets do a correlation plot to better compare the two and confirm it.

We can also do a paired t-test to compare the means of the two columns and be sure that there is no difference. To do it we use **Model** > **Hypothesis Tests**. The result it's showing at the table below.

```
       Paired t-test

data:  rain and prcp
t = 1.8, df = 3649, p-value = 0.08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02846  0.56142
sample estimates:
mean of the differences
                 0.2665
```

Figure 9: T-test between rain and prcp

The results are actually showing that there is defference between the two means. the P-value is 0.08 that is greater than 0.05, So we fail to accept the null hypothesis that is for no difference.

2.3 Correlation plot

To do it in R-instat we use **Describe** > **Multivariate** > **Correlations**, enter prcp and rain; then we click on Options and choose Scatter Matrix. The result is showing below.

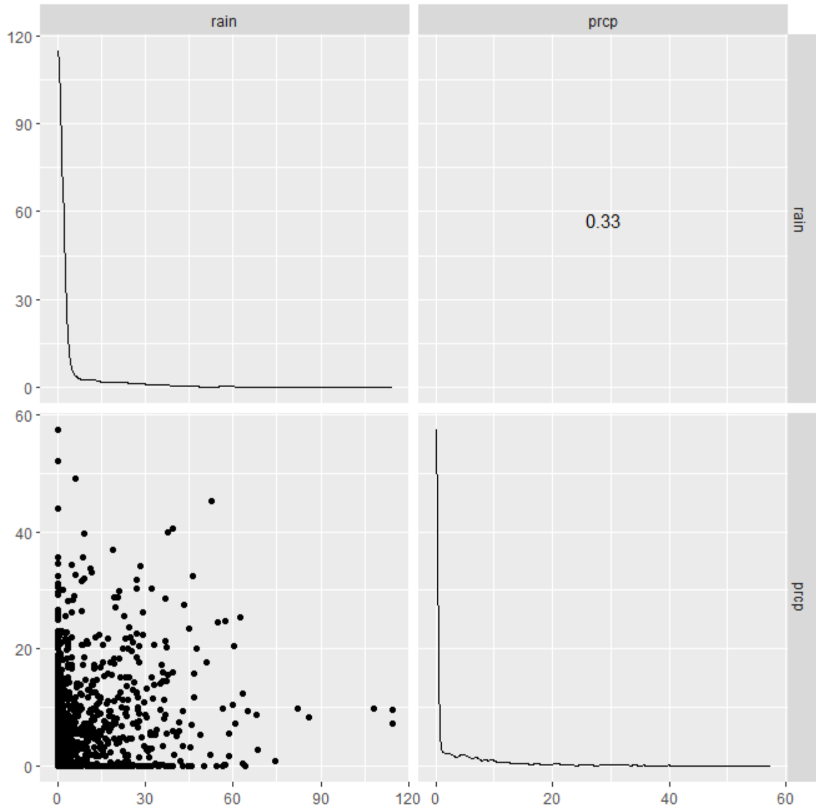

Figure 10: Correlation plot between Rain and Precipitation

Looking at this result, it seems like there is no difference regarding the shape of the satellite (prcp) and station data (rain) but the correlation is coefficient is not satisfied , 0.33 that means there is no correlation between the two. Lets look at the monthly correlation to have more precision.

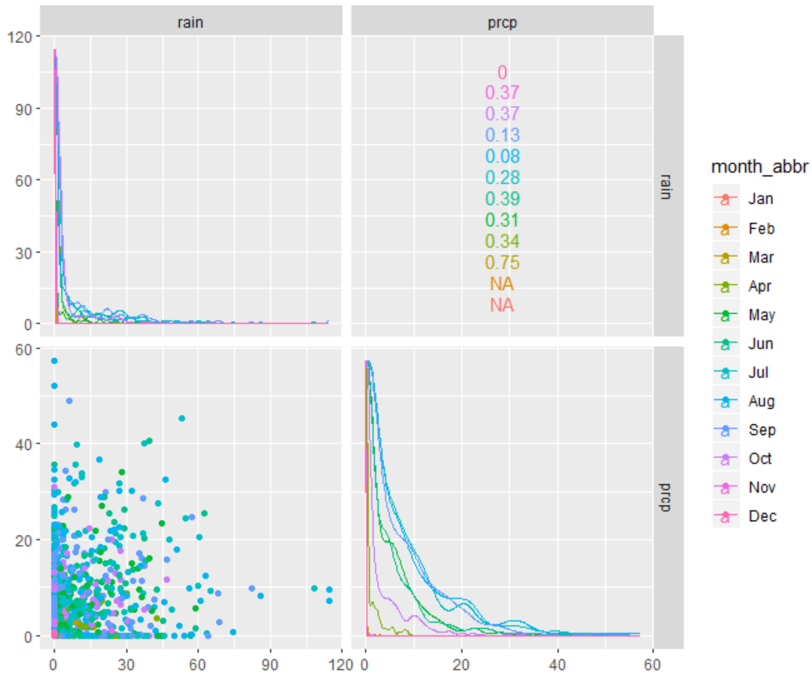

Figure 11: Correlation plot between Rain and Precipitation

Through this graph we can clearly see a difference between the shape of the satellite (prcp) and station data (rain). The shape differ from one month to another and the correlations are still small. It is just for October that we record a high correlation 0.75. This is clearly showing that shape of both variables depends on the month

Let's Fit a linear model to check whether month is significant. To do it we use **Model** > **Fit Model** > **General**. We obtain the following results:

```
Analysis of Variance Table

Response: rain
              Df Sum Sq Mean Sq F value Pr(>F)
month_abbr    11  30996    2818      37 <2e-16
Residuals   3638 277032      76
```

```
Analysis of Variance Table

Response: prcp
              Df Sum Sq Mean Sq F value Pr(>F)
month_abbr    11  25068    2279    88.5 <2e-16
Residuals   3641  93814      26
```

Figure 12: Anova table

For both of them (rain and prcp), variable month is significant. The p-value is less than 0.05. So the shape will actually depend of the month.

So let's look at the correlation plot month by month to have a better vew. To do it we use **Describe** > **Specific** > **Scatterplot** we complete the dialogue box and press on Plot Options and include the months as facets, We obtain the graph below

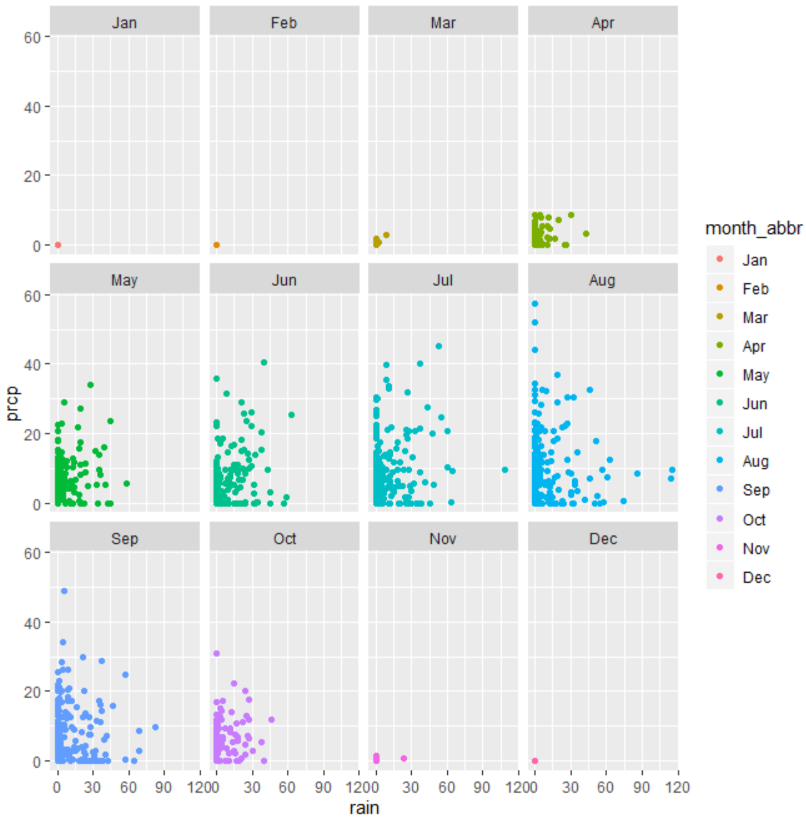

Figure 13: Monthly Correlation plot between Rain and Precipitation

As we see on the graph, January, February, March, November and December are actually dry month. May, June, July, August, September and October are the month were we usually have rain.

2.4 Markov chain with R

A Markov chain or Markov process, named after Russian mathematician, Andrey Markov (Shannon, 1948), is a mathematical system that undergoes transitions from one state to another (from a finite or countable number of possible states) in a chain like manner. It is a random process endowed with the Markov property. The Markov property states that the conditional probability distribution for a system at the next step given its current state depends only on the current state of the system, and not additionally on the state of the system at previous steps. Since the system changes randomly, it is generally impossible to predict the exact state of the system in the future.

For this part we will continue with our Garoua station data as our Rainfall dataset.The mains objectives will be to:

- Fit models to the chance of rain
- Find the best models using AIC and BIC analysis
- Plot the different models obtained.

2.4.1 Fit models to the chance of rain

Preliminary

We first load our data and remove the missing value.

```
station <- read.csv("garoua.csv")
station<-na.omit(station)
```

To predict the rainfall we will pay attention on two states of Markov chain dry "d" and wet "w". To specified it in R, we create a new variable call rainday that will record d if there was not rain and w if there was rain. This the code in R:

```
station$rainday[station$rain > 0.85]= "w"
station$rainday[station$rain <= 0.85]= "d"
```

We will also need lag variables that will record the value for the previous day.

```
station$lag1 <- lag(station$rainday,1)
station$lag2 <- lag(station$rainday,2)
station$lag3 <- lag(station$rainday,3)
```

Then we combine the lags to have the previous 2 days i.e. ”ww”, ”wd”, ”dw” or ”dd”; And the previous 3 days.
We can now start to fit our different models.

```
station$prev1 <- station$lag1
station$prev2 <- paste0(station$lag2, station$lag1)
station$prev3 <- paste0(station$lag3, station$lag2, station$lag1)
```

Fitting zero-order MC with 3 harmonics

The first model to fit is the zero-order MC(Markov chain) with 3 harmonics.

```
form_zero <- rainday ~ (cos(doy * 1 * 2 * pi/366) + sin(doy * 1 * 2 * pi/366) +
cos(doy * 2 * 2 * pi/366) + sin(doy * 2 * 2 * pi/366)+
cos(doy * 3 * 2 * pi/366) + sin(doy* 3 * 2 * pi/366))
fit_zero <- glm(as.formula(form_zero), data = station, family = binomial, na.act
```

We now look at the Anova test

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: rainday

Terms added sequentially (first to last)


                          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                       2888     2880.8
cos(doy * 1 * 2 * pi/366)  1   519.87      2887     2360.9 < 2.2e-16 ***
sin(doy * 1 * 2 * pi/366)  1   188.54      2886     2172.3 < 2.2e-16 ***
cos(doy * 2 * 2 * pi/366)  1    78.95      2885     2093.4 < 2.2e-16 ***
sin(doy * 2 * 2 * pi/366)  1    27.64      2884     2065.8 1.461e-07 ***
cos(doy * 3 * 2 * pi/366)  1     2.73      2883     2063.0   0.09850 .
sin(doy * 3 * 2 * pi/366)  1     3.71      2882     2059.3   0.05425 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 14: Anova table for zero-order MC with 3 harmonic

The 3 harmonic is not significant. The p value is greater than 0.05. So there is no need to do the 4 harmonic. We can still check to be sure.

The following table below is given the result for the 4 harmonic.

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: rainday

Terms added sequentially (first to last)


                         Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                      2888     2880.8
cos(doy * 1 * 2 * pi/366)  1   519.87      2887     2360.9 < 2.2e-16 ***
sin(doy * 1 * 2 * pi/366)  1   188.54      2886     2172.3 < 2.2e-16 ***
cos(doy * 2 * 2 * pi/366)  1    78.95      2885     2093.4 < 2.2e-16 ***
sin(doy * 2 * 2 * pi/366)  1    27.64      2884     2065.8 1.461e-07 ***
cos(doy * 3 * 2 * pi/366)  1     2.73      2883     2063.0   0.09850 .
sin(doy * 3 * 2 * pi/366)  1     3.71      2882     2059.3   0.05425 .
cos(doy * 4 * 2 * pi/366)  1     2.19      2881     2057.1   0.13862
sin(doy * 4 * 2 * pi/366)  1     2.88      2880     2054.3   0.08989 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 15: Anova table with 4 harmonic

As we were saying, even the 4 harmonic is not significant. The p-value is greater than 0.05. Let's move to the first order MC.

Fitting first-order MC

The table below is given us the Anova test for the first order MC without iteration.

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: rainday

Terms added sequentially (first to last)


                         Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                      2887     2880.3
cos(doy * 1 * 2 * pi/366)  1   519.49      2886     2360.8 < 2.2e-16 ***
sin(doy * 1 * 2 * pi/366)  1   188.50      2885     2172.3 < 2.2e-16 ***
cos(doy * 2 * 2 * pi/366)  1    78.92      2884     2093.4 < 2.2e-16 ***
sin(doy * 2 * 2 * pi/366)  1    27.64      2883     2065.8 1.463e-07 ***
prev1                      1     3.74      2882     2062.0   0.05309 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 16: Anova table fit first order Markov Chain without iteration

We can also look at the fitted first order Markov Chain with iteration to see if the result of the previous day affect on the present one. The result for the Anova as showing below.

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: rainday

Terms added sequentially (first to last)


                                Df Deviance Resid. Df Resid. Dev
NULL                                             2887     2880.3
cos(doy * 1 * 2 * pi/366)        1   519.49      2886     2360.8
sin(doy * 1 * 2 * pi/366)        1   188.50      2885     2172.3
cos(doy * 2 * 2 * pi/366)        1    78.92      2884     2093.4
sin(doy * 2 * 2 * pi/366)        1    27.64      2883     2065.8
prev1                            1     3.74      2882     2062.0
cos(doy * 1 * 2 * pi/366):prev1  1     4.14      2881     2057.9
sin(doy * 1 * 2 * pi/366):prev1  1     5.83      2880     2052.1
cos(doy * 2 * 2 * pi/366):prev1  1     2.96      2879     2049.1
sin(doy * 2 * 2 * pi/366):prev1  1     0.42      2878     2048.7
                                 Pr(>Chi)
NULL
cos(doy * 1 * 2 * pi/366)       < 2.2e-16 ***
sin(doy * 1 * 2 * pi/366)       < 2.2e-16 ***
cos(doy * 2 * 2 * pi/366)       < 2.2e-16 ***
sin(doy * 2 * 2 * pi/366)       1.463e-07 ***
prev1                             0.05309 .
cos(doy * 1 * 2 * pi/366):prev1   0.04199 *
sin(doy * 1 * 2 * pi/366):prev1   0.01572 *
cos(doy * 2 * 2 * pi/366):prev1   0.08542 .
sin(doy * 2 * 2 * pi/366):prev1   0.51446
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 17: Anova table fit first order Markov Chain with iteration

Looking at this result, we can see that the result for the previous day is significant at 0.05 level of significant. That means the fact that the day before was dry for instance can affect the day after to be also dry. Our data can confirm it.

day	rain	tmin	tmax	rainday	lag1	lag2	lag3	prev1
32	0.0	19.0	37.3	d	NA	NA	NA	NA
33	0.0	21.3	37.5	d	d	NA	NA	d
34	0.0	20.0	36.0	d	d	d	NA	d
35	0.0	24.0	33.6	d	d	d	d	d
36	0.0	21.5	32.5	d	d	d	d	d
37	0.0	23.4	30.3	d	d	d	d	d
38	0.0	21.0	30.8	d	d	d	d	d
39	0.0	19.8	28.5	d	d	d	d	d
40	0.0	20.4	27.2	d	d	d	d	d
41	0.0	21.0	28.5	d	d	d	d	d
42	0.0	20.5	31.2	d	d	d	d	d
43	0.0	20.2	32.5	d	d	d	d	d
44	0.0	21.0	33.0	d	d	d	d	d
45	0.0	20.2	34.5	d	d	d	d	d
46	0.0	22.5	34.5	d	d	d	d	d

Figure 18: A vew of Garoua Rainfall data

Fitting second-order MC

The result for the Anova test is present at the table below

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: rainday

Terms added sequentially (first to last)


                          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                       2886     2879.9
cos(doy * 1 * 2 * pi/366)  1   519.10      2885     2360.8 < 2.2e-16 ***
sin(doy * 1 * 2 * pi/366)  1   188.46      2884     2172.3 < 2.2e-16 ***
cos(doy * 2 * 2 * pi/366)  1    78.90      2883     2093.4 < 2.2e-16 ***
sin(doy * 2 * 2 * pi/366)  1    27.63      2882     2065.8 1.465e-07 ***
prev2                      3     4.42      2879     2061.3    0.2193
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 19: Anova table fit second order Markov Chain without iteration

Without iteration , the variable **prev2** don't have effect on the rainday. let check if with iteration the value of the 2 previous day should have an impact on the actually value.
The result for the second order MC with iteration is present on the table below.

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: rainday

Terms added sequentially (first to last)


                              Df Deviance Resid. Df Resid. Dev
NULL                                           2886     2879.9
cos(doy * 1 * 2 * pi/366)      1   519.10      2885     2360.8
sin(doy * 1 * 2 * pi/366)      1   188.46      2884     2172.3
cos(doy * 2 * 2 * pi/366)      1    78.90      2883     2093.4
sin(doy * 2 * 2 * pi/366)      1    27.63      2882     2065.8
prev2                          3     4.42      2879     2061.3
cos(doy * 1 * 2 * pi/366):prev2  3    14.90      2876     2046.5
sin(doy * 1 * 2 * pi/366):prev2  3    13.84      2873     2032.6
cos(doy * 2 * 2 * pi/366):prev2  3     2.90      2870     2029.7
sin(doy * 2 * 2 * pi/366):prev2  3     3.69      2867     2026.0
                                Pr(>Chi)
NULL
cos(doy * 1 * 2 * pi/366)      < 2.2e-16 ***
sin(doy * 1 * 2 * pi/366)      < 2.2e-16 ***
cos(doy * 2 * 2 * pi/366)      < 2.2e-16 ***
sin(doy * 2 * 2 * pi/366)      1.465e-07 ***
prev2                           0.219337
cos(doy * 1 * 2 * pi/366):prev2  0.001907 **
sin(doy * 1 * 2 * pi/366):prev2  0.003135 **
cos(doy * 2 * 2 * pi/366):prev2  0.406568
sin(doy * 2 * 2 * pi/366):prev2  0.297320
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 20: Anova table fit second order Markov Chain with iteration

We can see that the prev2 is highly significant at level 0.01. So the value of the 2 previous day will have an impact on the future day.
Our last model is the third order. So let have a look on it.

Fitting third-order MC

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: rainday

Terms added sequentially (first to last)


                        Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                     2885     2879.4
cos(doy * 1 * 2 * pi/366)  1   518.72      2884     2360.7 < 2.2e-16 ***
sin(doy * 1 * 2 * pi/366)  1   188.43      2883     2172.3 < 2.2e-16 ***
cos(doy * 2 * 2 * pi/366)  1    78.88      2882     2093.4 < 2.2e-16 ***
sin(doy * 2 * 2 * pi/366)  1    27.63      2881     2065.8 1.467e-07 ***
prev3                      7     9.33      2874     2056.4    0.2301
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 21: Anova table fit third order Markov Chain without iteration

Without iteration ,also the variable **prev3** don't have effect on the rainday. let check if with iteration the value of the 3 previous day should have an impact on the actually value.
The result for the third order MC with iteration is present on the table below.

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: rainday

Terms added sequentially (first to last)


                              Df Deviance Resid. Df Resid. Dev
NULL                                           2885     2879.4
cos(doy * 1 * 2 * pi/366)        1   518.72      2884     2360.7
sin(doy * 1 * 2 * pi/366)        1   188.43      2883     2172.3
cos(doy * 2 * 2 * pi/366)        1    78.88      2882     2093.4
sin(doy * 2 * 2 * pi/366)        1    27.63      2881     2065.8
prev3                            7     9.33      2874     2056.4
cos(doy * 1 * 2 * pi/366):prev3  7    22.44      2867     2034.0
sin(doy * 1 * 2 * pi/366):prev3  7    23.17      2860     2010.8
cos(doy * 2 * 2 * pi/366):prev3  7    10.21      2853     2000.6
sin(doy * 2 * 2 * pi/366):prev3  7     8.89      2846     1991.7
                                 Pr(>Chi)
NULL
cos(doy * 1 * 2 * pi/366)       < 2.2e-16 ***
sin(doy * 1 * 2 * pi/366)       < 2.2e-16 ***
cos(doy * 2 * 2 * pi/366)       < 2.2e-16 ***
sin(doy * 2 * 2 * pi/366)       1.467e-07 ***
prev3                            0.230087
cos(doy * 1 * 2 * pi/366):prev3  0.002130 **
sin(doy * 1 * 2 * pi/366):prev3  0.001591 **
cos(doy * 2 * 2 * pi/366):prev3  0.177234
sin(doy * 2 * 2 * pi/366):prev3  0.260734
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 22: Anova table fit third order Markov Chain with iteration

We can see that the **prev3** is highly significant at level 0.01. So the value of the 2 previous day will have an impact on the future day.
Now lets see which of this models are better. To do it we need to check the AIC and BIC.

2.4.2 AIC and BIC Analysis

To do it, we type the following code:

```
AIC(fit_zero, fit_one, fit_one_int, fit_two, fit_two_int, fit_three, fit_three_i
BIC(fit_zero, fit_one, fit_one_int, fit_two, fit_two_int, fit_three, fit_three_i
```

The results are present below

	df	AIC
fit_zero	7	2073.330
fit_one	6	2074.024
fit_one_int	10	2068.671
fit_two	8	2077.342
fit_two_int	20	2066.017
fit_three	12	2080.438
fit_three_int	40	2071.729

	df	BIC
fit_zero	7	2115.111
fit_one	6	2109.834
fit_one_int	10	2128.354
fit_two	8	2125.086
fit_two_int	20	2185.376
fit_three	12	2152.050
fit_three_int	40	2310.434

Figure 23: AIC and BIC analysis

Looking at the AIC **fit_two_int** is the better one. Looking at the BIC **fit_one** is the better one. Since fit one have the smaller degree of freedom, it can be consider as the better one.

2.4.3 Plots visualization

Plotting zero-order MC model

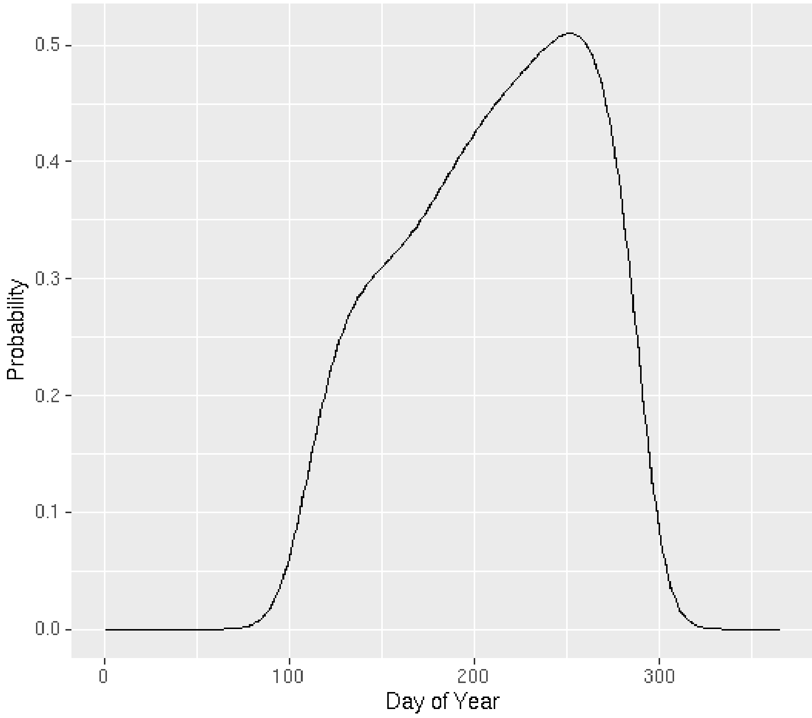

Figure 24: Plot of zero-order MC model

Plotting first-order MC model

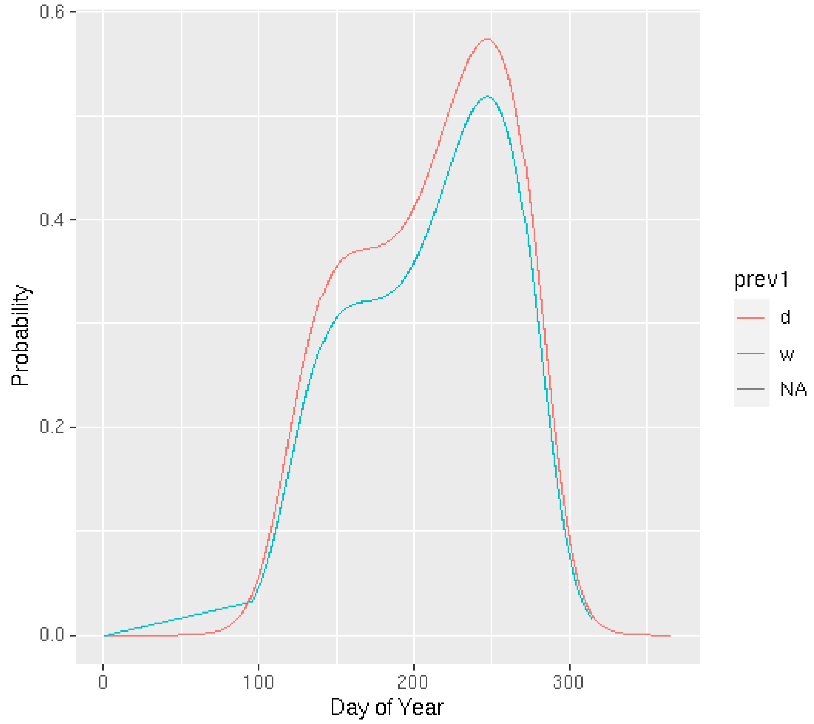

Figure 25: Plot of first-order MC model

Looking at the graph we can see that the probability of having a rain day when the previous day is dry is higher than when the previous day is wet. have rainy day.

Plotting first-order MC model with iteration

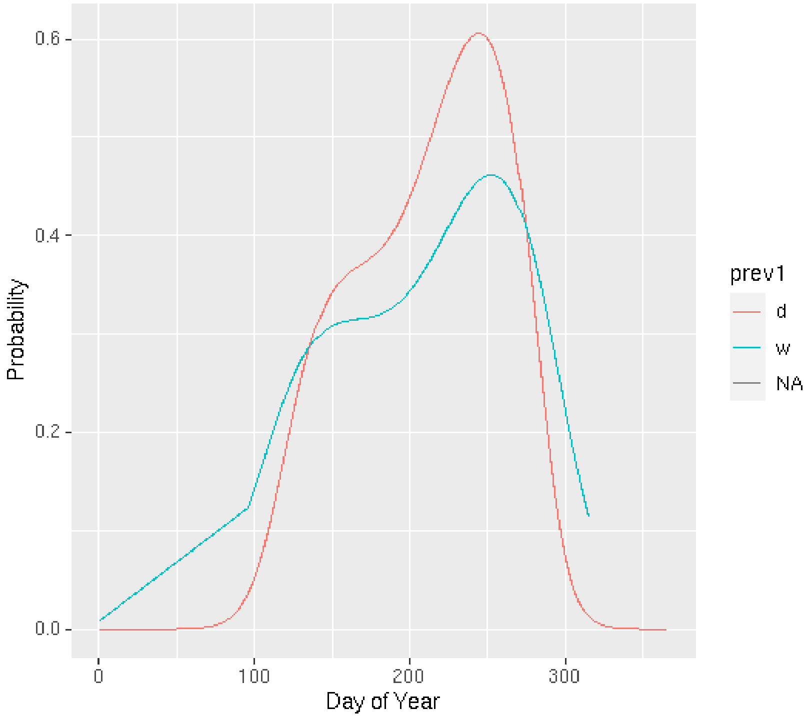

Figure 26: Plot of first-order MC model with iteration

Plotting second-order MC model without iteration

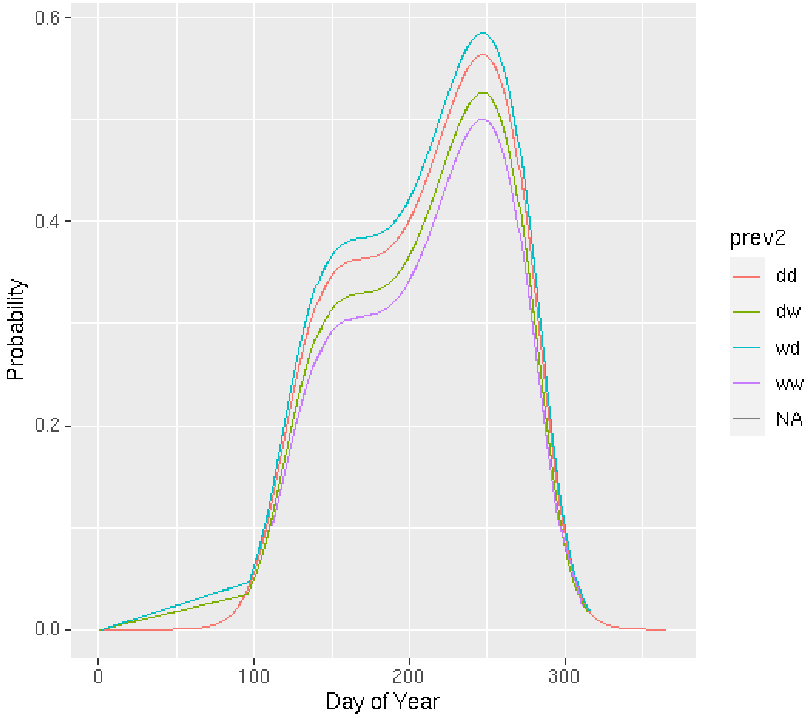

Figure 27: Plot of second-order MC model without iteration

The probability of have rain day is more higher when the two previous day was wet and dry.

Plotting third-order MC model without iteration

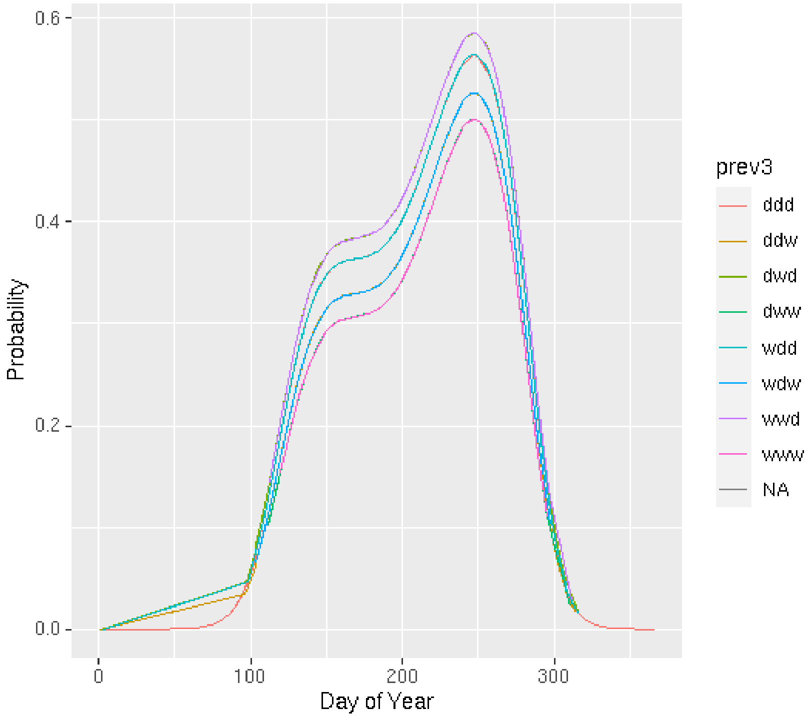

Figure 28: Plot of third-order MC model without iteration

The probability of have rain day is more higher when the three previous day was wet, wet and dry.