

FirstRun

Lezheng Fang

2/13/2022

Read Climate and Cherry Blossom Data

```
library(Amelia)
```

```
## Loading required package: Rcpp
```

```
## ##
```

```
## ## Amelia II: Multiple Imputation
```

```
## ## (Version 1.7.6, built: 2019-11-24)
```

```
## ## Copyright (C) 2005-2022 James Honaker, Gary King and Matthew Blackwell
```

```
## ## Refer to http://gking.harvard.edu/amelia/ for more information
```

```
## ##
```

```
climate_name = "../data/cleaned_with_NA_kyoto.csv"
```

```
blossom_name = "../data/kyoto.csv"
```

```
climate <- read.csv(climate_name)
```

```
blossom <- read.csv(blossom_name)
```

```
summary(blossom)
```

```
##   location      lat      long      alt      year
## kyoto:833  Min.   :35.01  Min.   :135.7  Min.   :44  Min.   : 812
##           1st Qu.:35.01  1st Qu.:135.7  1st Qu.:44  1st Qu.:1329
##           Median :35.01  Median :135.7  Median :44  Median :1586
##           Mean   :35.01  Mean   :135.7  Mean   :44  Mean   :1552
##           3rd Qu.:35.01  3rd Qu.:135.7  3rd Qu.:44  3rd Qu.:1808
##           Max.   :35.01  Max.   :135.7  Max.   :44  Max.   :2021
```

```
##
```

```
##      bloom_date  bloom_doy
## 0812-04-01: 1  Min.   : 85.0
## 0815-04-15: 1  1st Qu.:100.0
## 0831-04-06: 1  Median :104.0
## 0851-04-18: 1  Mean   :104.5
## 0853-04-14: 1  3rd Qu.:109.0
## 0864-04-09: 1  Max.   :124.0
## (Other)    :827
```

```
str(blossom)
```

```
## 'data.frame': 833 obs. of 7 variables:
```

```
## $ location : Factor w/ 1 level "kyoto": 1 1 1 1 1 1 1 1 1 1 ...
```

```
## $ lat : num 35 35 35 35 35 ...
```

```
## $ long : num 136 136 136 136 136 ...
```

```
## $ alt : int 44 44 44 44 44 44 44 44 44 44 ...
```

```
## $ year      : int  812 815 831 851 853 864 866 869 889 891 ...
## $ bloom_date: Factor w/ 833 levels "0812-04-01","0815-04-15",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ bloom_doy : int  92 105 96 108 104 100 106 95 104 109 ...
```

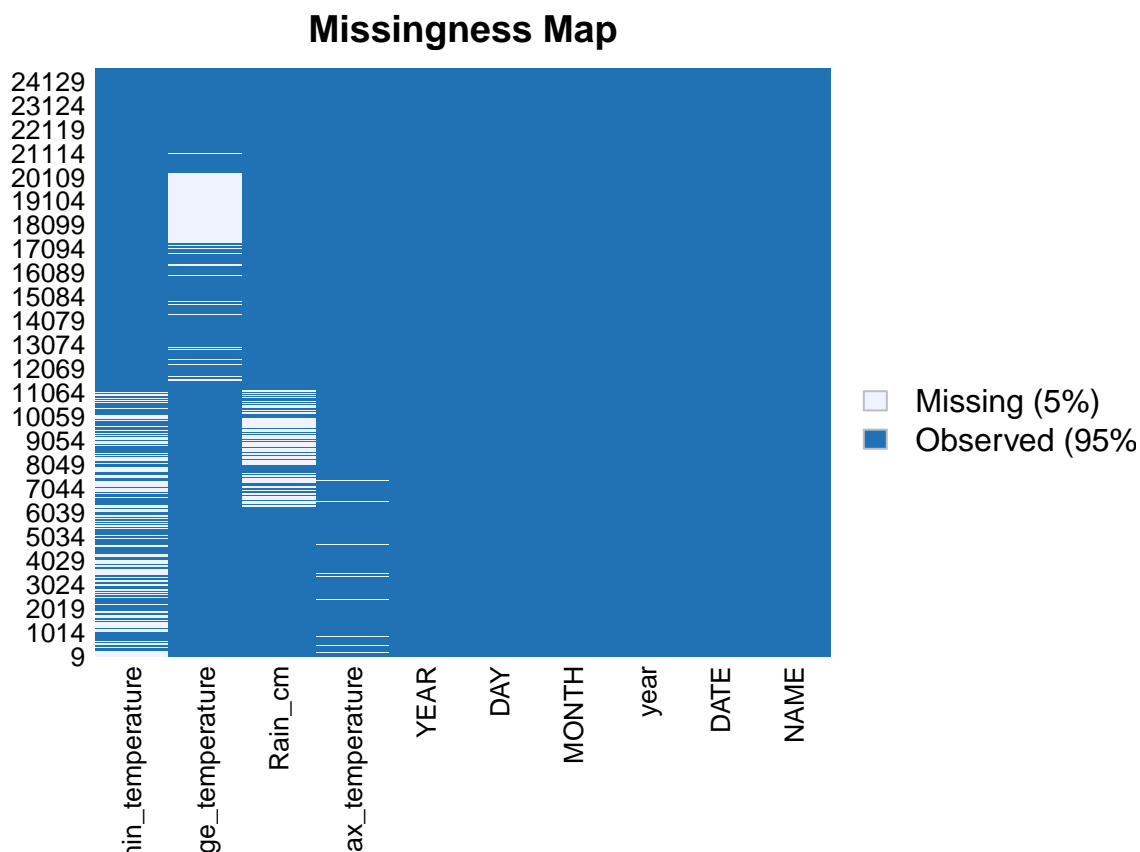
```
summary(climate)
```

```
##          NAME          DATE          Rain_cm          average_temperature
## KYOTO, JA:24654  1953-02-01:    1  Min.    : 0.000  Min.    : -3.0
##                  1953-02-02:    1  1st Qu.: 0.000  1st Qu.:  7.9
##                  1953-02-03:    1  Median : 0.000  Median :16.2
##                  1953-02-04:    1  Mean     : 4.852  Mean     :15.9
##                  1953-02-05:    1  3rd Qu.: 2.500  3rd Qu.:23.4
##                  1953-02-06:    1  Max.      :288.600  Max.      :32.9
##                  (Other)    :24648  NA's      :3228    NA's      :3706
## max_temperature min_temperature      year      MONTH
## Min.    : 0.3    Min.    : -8.10  Min.    :1953  Min.    : 1.000
## 1st Qu.:12.8    1st Qu.:  3.10  1st Qu.:1969  1st Qu.: 4.000
## Median :21.2    Median :11.10  Median :1986  Median : 7.000
## Mean     :20.8    Mean     :11.29  Mean     :1987  Mean     : 6.513
## 3rd Qu.:28.3    3rd Qu.:19.20  3rd Qu.:2004  3rd Qu.:10.000
## Max.      :39.8    Max.      :28.70  Max.      :2022  Max.      :12.000
## NA's      :459    NA's      :5088
##          DAY          YEAR
## Min.    : 1.00    Min.    :1953
## 1st Qu.: 8.00    1st Qu.:1969
## Median :16.00    Median :1986
## Mean     :15.71    Mean     :1987
## 3rd Qu.:23.00    3rd Qu.:2004
## Max.      :31.00    Max.      :2022
##
```

```
str(climate)
```

```
## 'data.frame':    24654 obs. of  10 variables:
## $ NAME          : Factor w/ 1 level "KYOTO, JA": 1 1 1 1 1 1 1 1 1 1 ...
## $ DATE          : Factor w/ 24654 levels "1953-02-01","1953-02-02",...: 1 2 3 4 5 6 7 8 9 10 ..
## $ Rain_cm       : num  0.4 0.1 0 0 0 0 0 0 0 5.6 ...
## $ average_temperature: num  0.6 1.4 2.1 2.2 2.7 2.8 3.3 1.2 3.1 6.4 ...
## $ max_temperature  : num  4.9 6.2 7.4 4.7 8.4 6.6 10.1 7.3 8.7 10.8 ...
## $ min_temperature  : num  -2.4 -3.5 -1.4 -1.4 0.4 -3.2 -0.4 -3.5 -3.4 0.6 ...
## $ year            : int  1953 1953 1953 1953 1953 1953 1953 1953 1953 1953 ...
## $ MONTH           : int   2 2 2 2 2 2 2 2 2 ...
## $ DAY             : int   1 2 3 4 5 6 7 8 9 10 ...
## $ YEAR            : int  1953 1953 1953 1953 1953 1953 1953 1953 1953 1953 ...
```

```
missmap(climate)
```



Impute Missing data in climate

```
set.seed(1)
m = 5
climate$doy = as.Date(climate$DATE)
str(climate$doy)
```

```
## Date[1:24654], format: "1953-02-01" "1953-02-02" "1953-02-03" "1953-02-04" "1953-02-05" ...
climate_amelia <- amelia(x = climate, idvars = c("NAME", "MONTH", "DAY", "YEAR", "DATE", "year"), ts = "doy")
```

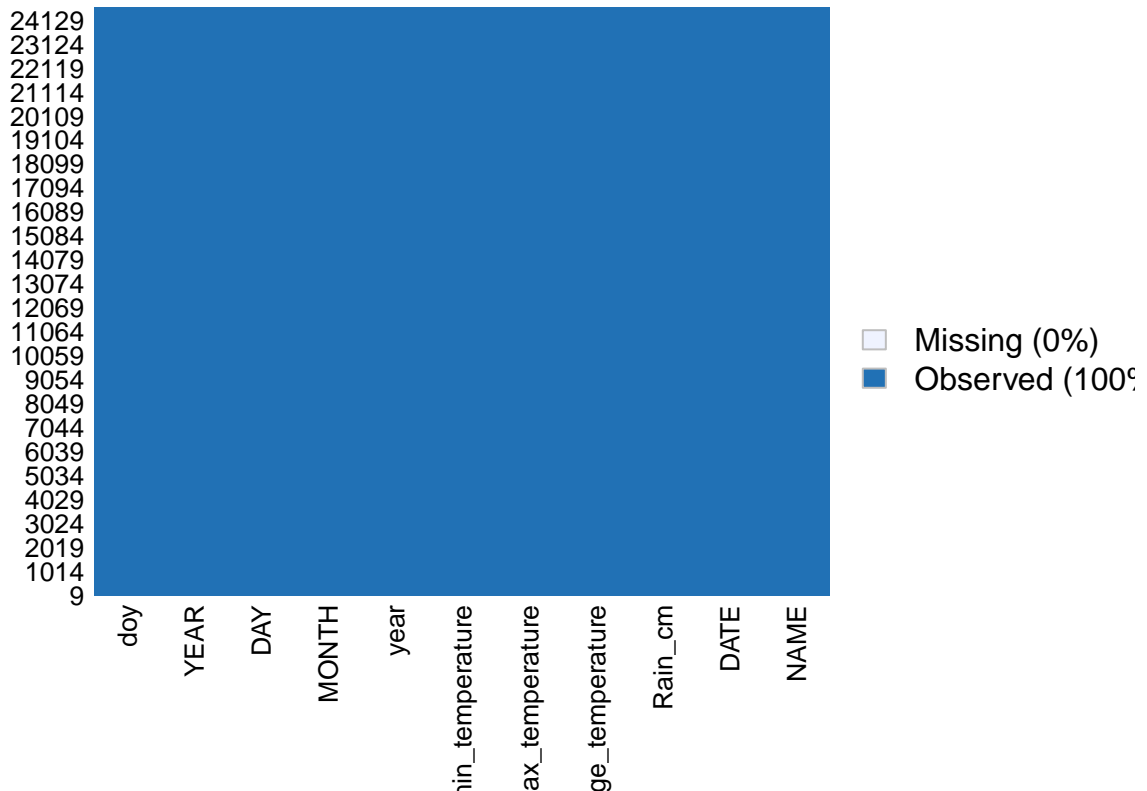
```
## -- Imputation 1 --
##
## 1 2 3 4 5
##
## -- Imputation 2 --
##
## 1 2 3 4 5
##
## -- Imputation 3 --
##
## 1 2 3 4 5
##
## -- Imputation 4 --
##
## 1 2 3 4 5
```

```
##
## -- Imputation 5 --
##
## 1 2 3 4 5
str(climate_amelia$imputations$imp1)

## 'data.frame': 24654 obs. of 11 variables:
## $ NAME : Factor w/ 1 level "KYOTO, JA": 1 1 1 1 1 1 1 1 1 1 ...
## $ DATE : Factor w/ 24654 levels "1953-02-01","1953-02-02",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Rain_cm : num 0.4 0.1 0 0 0 0 0 0 0 5.6 ...
## $ average_temperature: num 0.6 1.4 2.1 2.2 2.7 2.8 3.3 1.2 3.1 6.4 ...
## $ max_temperature : num 4.9 6.2 7.4 4.7 8.4 6.6 10.1 7.3 8.7 10.8 ...
## $ min_temperature : num -2.4 -3.5 -1.4 -1.4 0.4 -3.2 -0.4 -3.5 -3.4 0.6 ...
## $ year : int 1953 1953 1953 1953 1953 1953 1953 1953 1953 1953 ...
## $ MONTH : int 2 2 2 2 2 2 2 2 2 2 ...
## $ DAY : int 1 2 3 4 5 6 7 8 9 10 ...
## $ YEAR : int 1953 1953 1953 1953 1953 1953 1953 1953 1953 1953 ...
## $ doy : Date, format: "1953-02-01" "1953-02-02" ...

climate_imputed <- climate
# Average the imputations between different simulated datasets
col_index = which(names(climate_amelia$imputations$imp1) %in% c("average_temperature", "max_tempera
for( col in col_index){
  temp2=numeric()
  for (i in 1:m){
    temp2 = cbind(temp2, climate_amelia$imputations[[i]][,col])
  }
  climate_imputed[,col] = apply(temp2, 1, mean)
}
missmap(climate_imputed)
```

Missingness Map



```
head(climate_imputed)
```

```
##      NAME      DATE Rain_cm average_temperature max_temperature
## 1 KYOTO, JA 1953-02-01      0.4                0.6              4.9
## 2 KYOTO, JA 1953-02-02      0.1                1.4              6.2
## 3 KYOTO, JA 1953-02-03      0.0                2.1              7.4
## 4 KYOTO, JA 1953-02-04      0.0                2.2              4.7
## 5 KYOTO, JA 1953-02-05      0.0                2.7              8.4
## 6 KYOTO, JA 1953-02-06      0.0                2.8              6.6
## min_temperature year MONTH DAY YEAR      doy
## 1          -2.4 1953      2   1 1953 1953-02-01
## 2          -3.5 1953      2   2 1953 1953-02-02
## 3          -1.4 1953      2   3 1953 1953-02-03
## 4          -1.4 1953      2   4 1953 1953-02-04
## 5           0.4 1953      2   5 1953 1953-02-05
## 6          -3.2 1953      2   6 1953 1953-02-06
```

```
# Saving imputed Data
```

```
#saveRDS(climate_imputed,file = "climate_imputed.csv")
```

```
write.table(climate_imputed,'../data/kyoto_climate_imputed.csv',sep="," ,col.names = c(names(climate_imputed)),
            as.is=TRUE)
climate_imputed = read.table('../data/kyoto_climate_imputed.csv' ,sep="," ,header=TRUE)
climate_imputed$DATE = as.Date(climate_imputed$DATE )
climate_imputed$doy = NULL
str(climate_imputed)
```

```
## 'data.frame':   24654 obs. of  10 variables:
```

```
## $ NAME           : Factor w/ 1 level "KYOTO, JA": 1 1 1 1 1 1 1 1 1 1 ...
## $ DATE           : Date, format: "1953-02-01" "1953-02-02" ...
## $ Rain_cm        : num  0.4 0.1 0 0 0 0 0 0 0 5.6 ...
## $ average_temperature: num  0.6 1.4 2.1 2.2 2.7 2.8 3.3 1.2 3.1 6.4 ...
## $ max_temperature  : num  4.9 6.2 7.4 4.7 8.4 6.6 10.1 7.3 8.7 10.8 ...
## $ min_temperature  : num  -2.4 -3.5 -1.4 -1.4 0.4 -3.2 -0.4 -3.5 -3.4 0.6 ...
## $ year            : int   1953 1953 1953 1953 1953 1953 1953 1953 1953 1953 ...
## $ MONTH           : int    2 2 2 2 2 2 2 2 2 2 ...
## $ DAY             : int    1 2 3 4 5 6 7 8 9 10 ...
## $ YEAR            : int   1953 1953 1953 1953 1953 1953 1953 1953 1953 1953 ...
```

```
#readRDS()
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this: