Exploratory Analysis of NOAA Weather Data

Benj, Grace and Mark

Outline

Introduction

Dataset

Data Cleaning

Variance

NA Replacement

Clustering

Conclusion

Dataset

Dataset

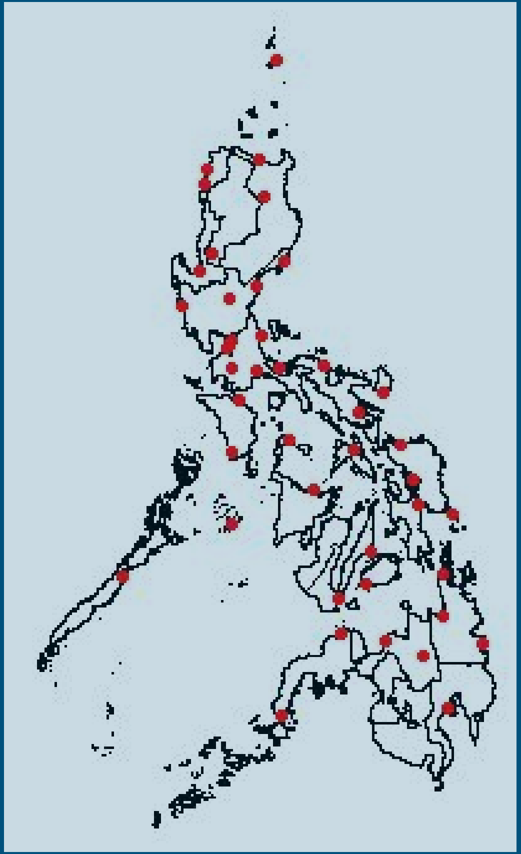
Daily temperature (°C) and amount of precipitation (mm)

88 weather stations across the Philippines

January 1, 1960 to June 21, 2015

20261 rows and 89 columns

Longitude and latitude of the stations



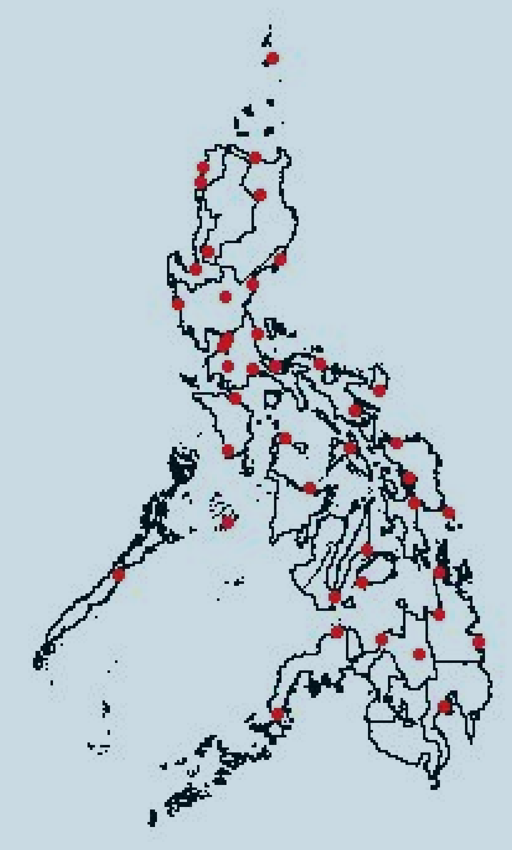
Dataset

National Oceanic and Atmospheric Administration's (NOAA) Integrated Surface Data (ISD)

ftp://ftp.ncdc.noaa.gov/pub/data/noaa/

1901 to 2016

293 Countries

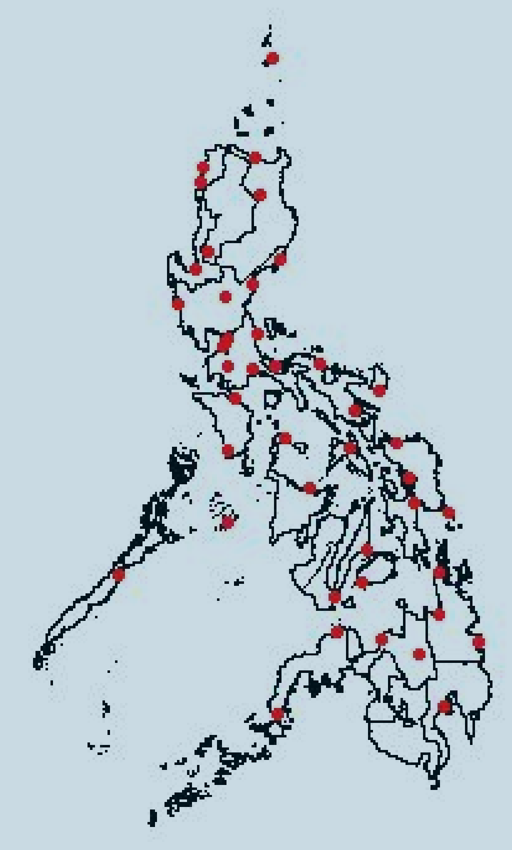


Data Cleaning

Data Cleaning

88 weather stations across the Philippines but reduced to 85 stations

- Baler + Baler Radar
- Basco + Basco Radar
- Davao + Davao Airport



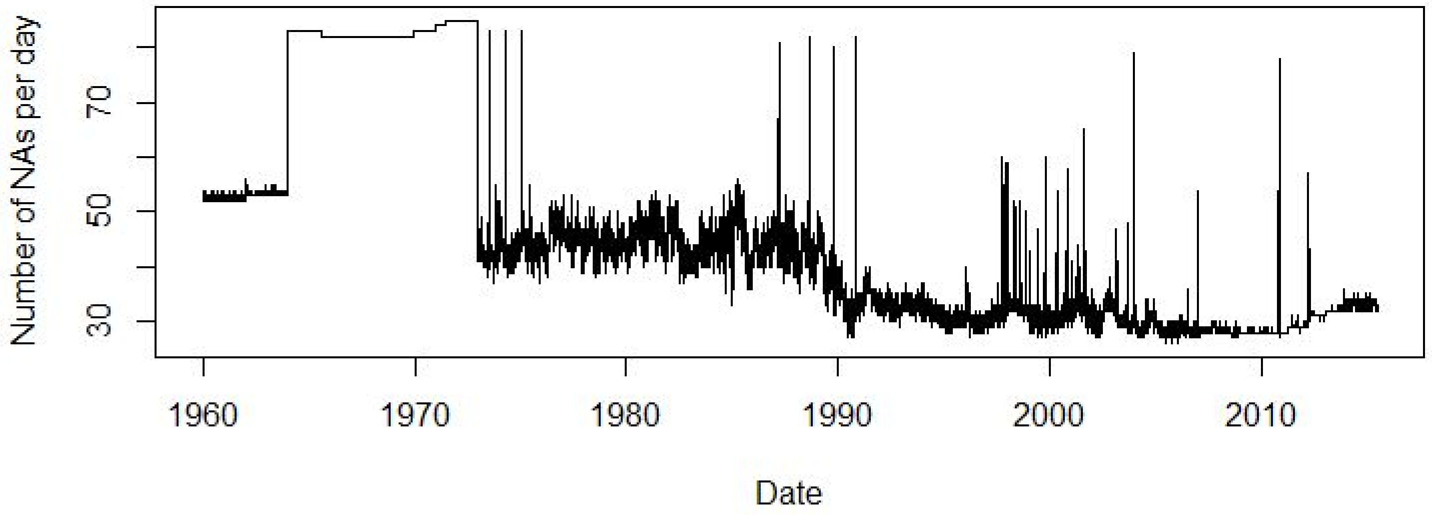
Missing Values

There’s a lot of missing (NA) values in the dataset probably caused by

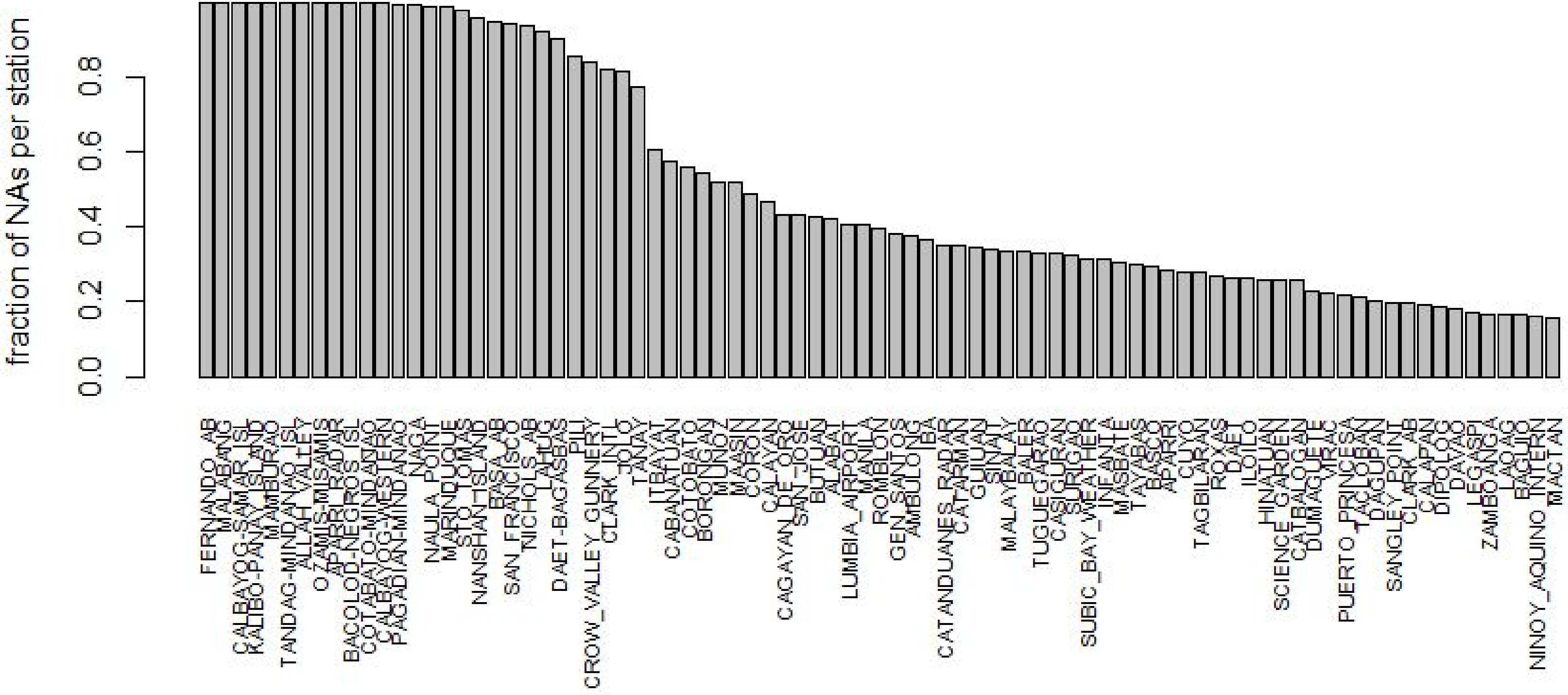
- Station is not yet established
- Station experienced difficulties in reading data
- Etc

The following plots are:

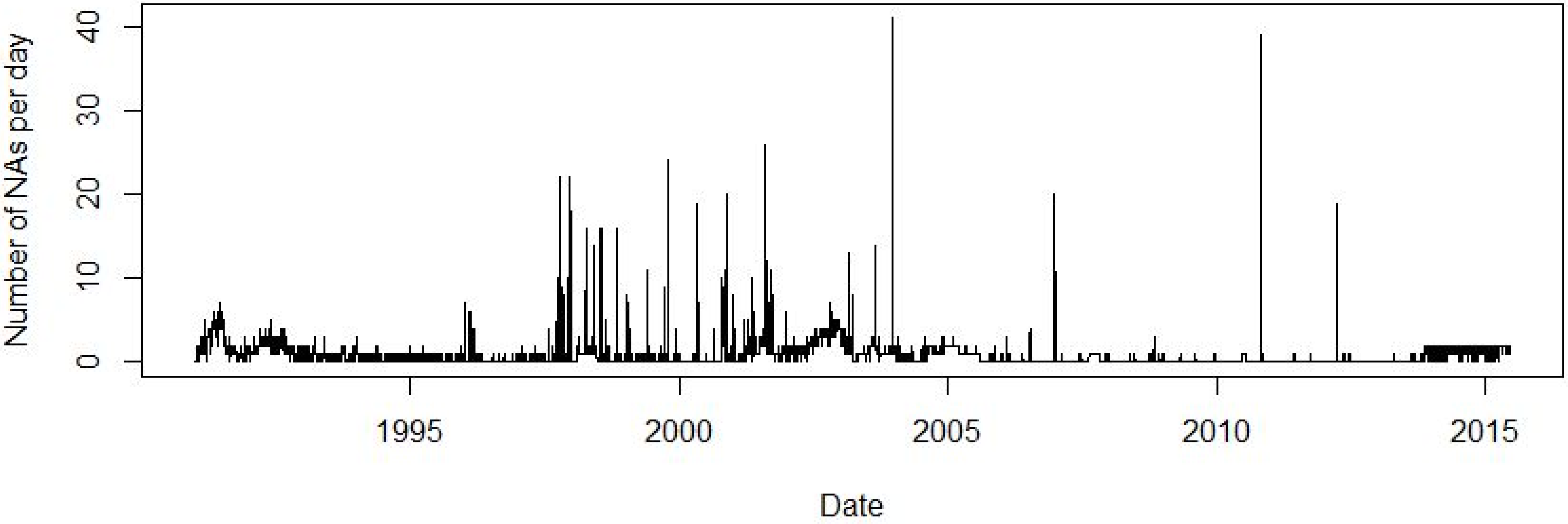
- # of NAs over a period of time
- Distribution of NAs among the stations



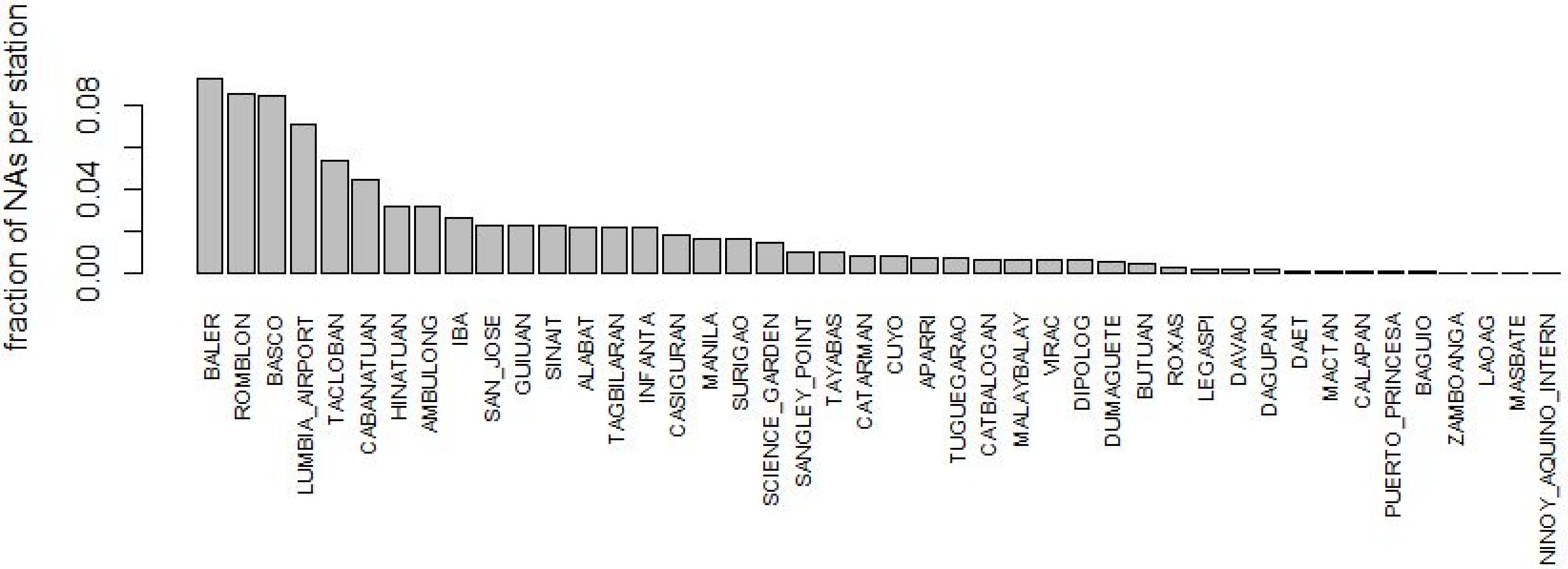
A lot of missing data between 1966 to 1974



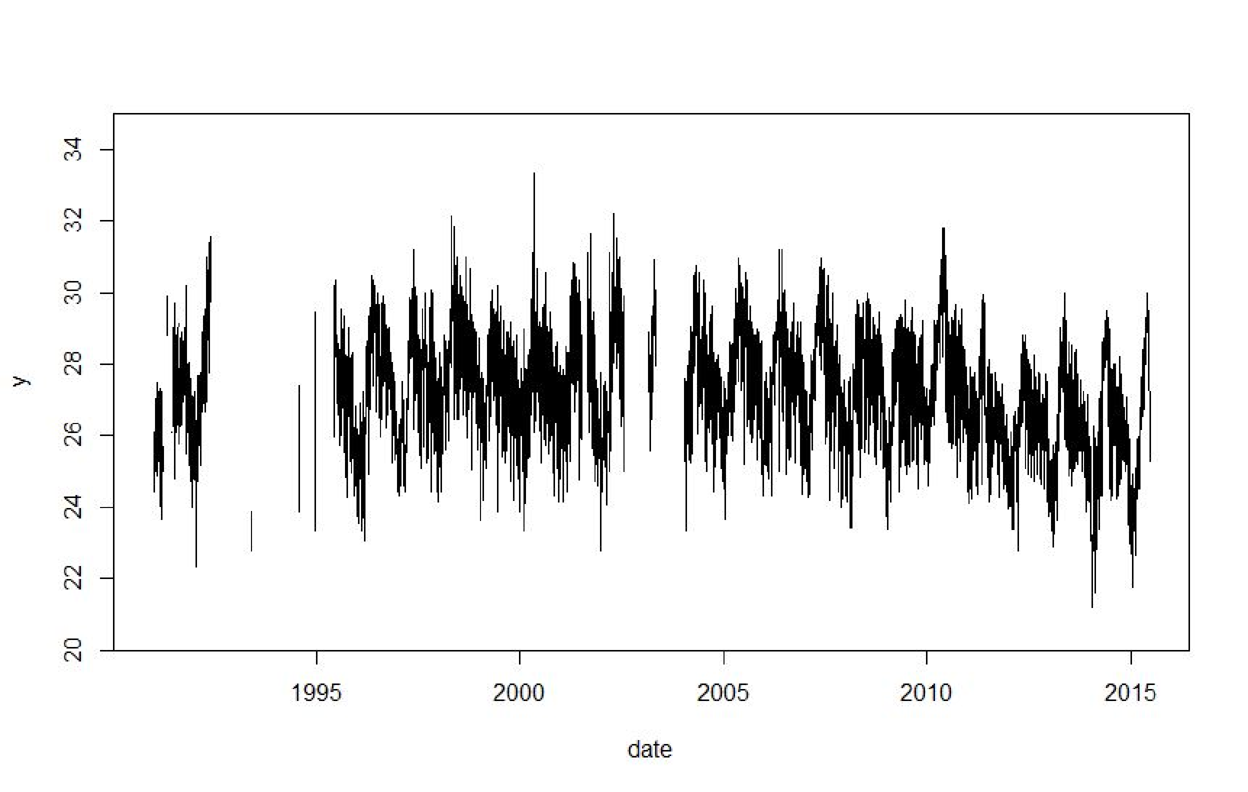
Stations with the most number of NAs



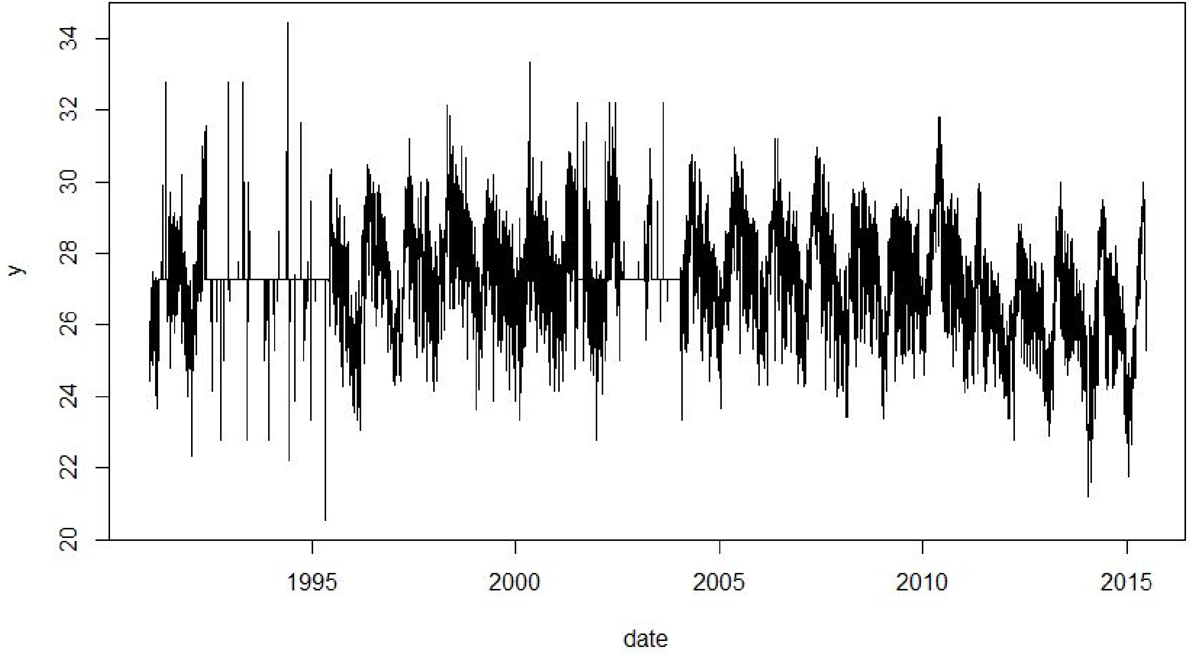
We limit the timeframe between 1990 to 2015



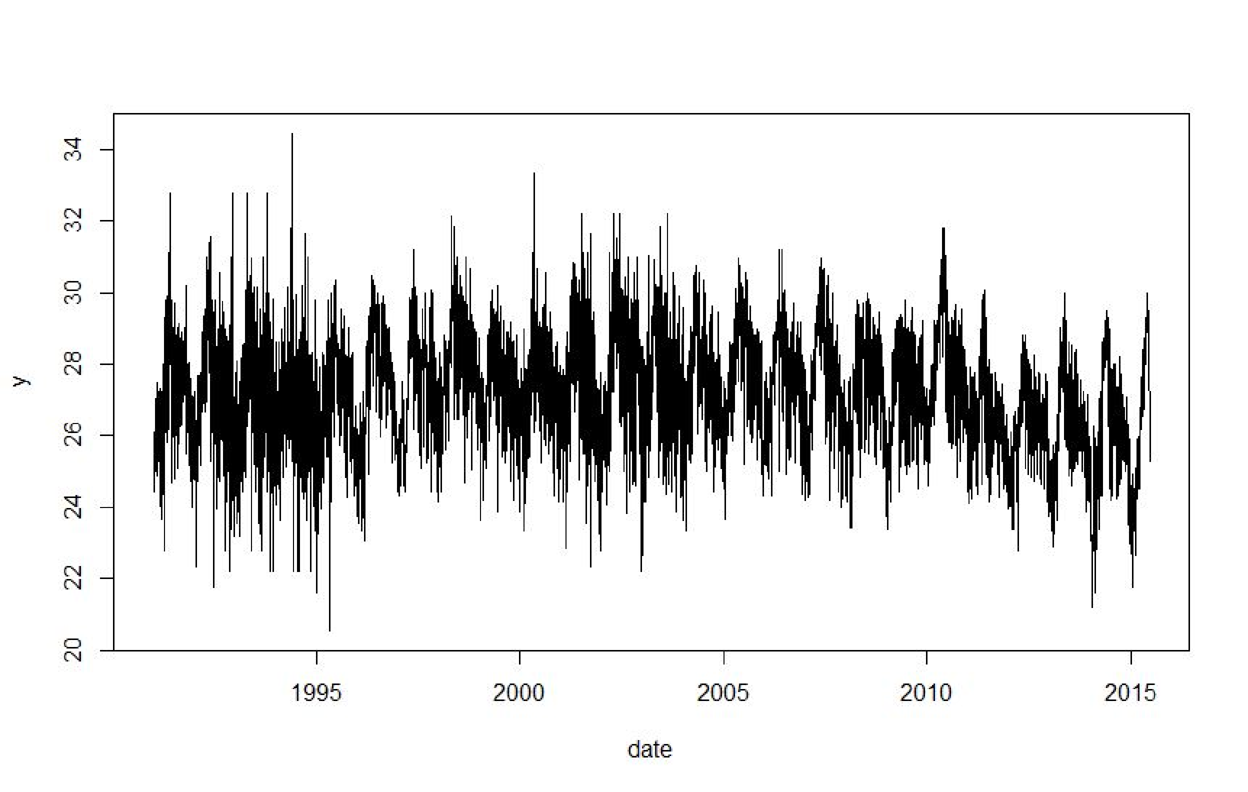
Removed some of the stations



Data with missing values



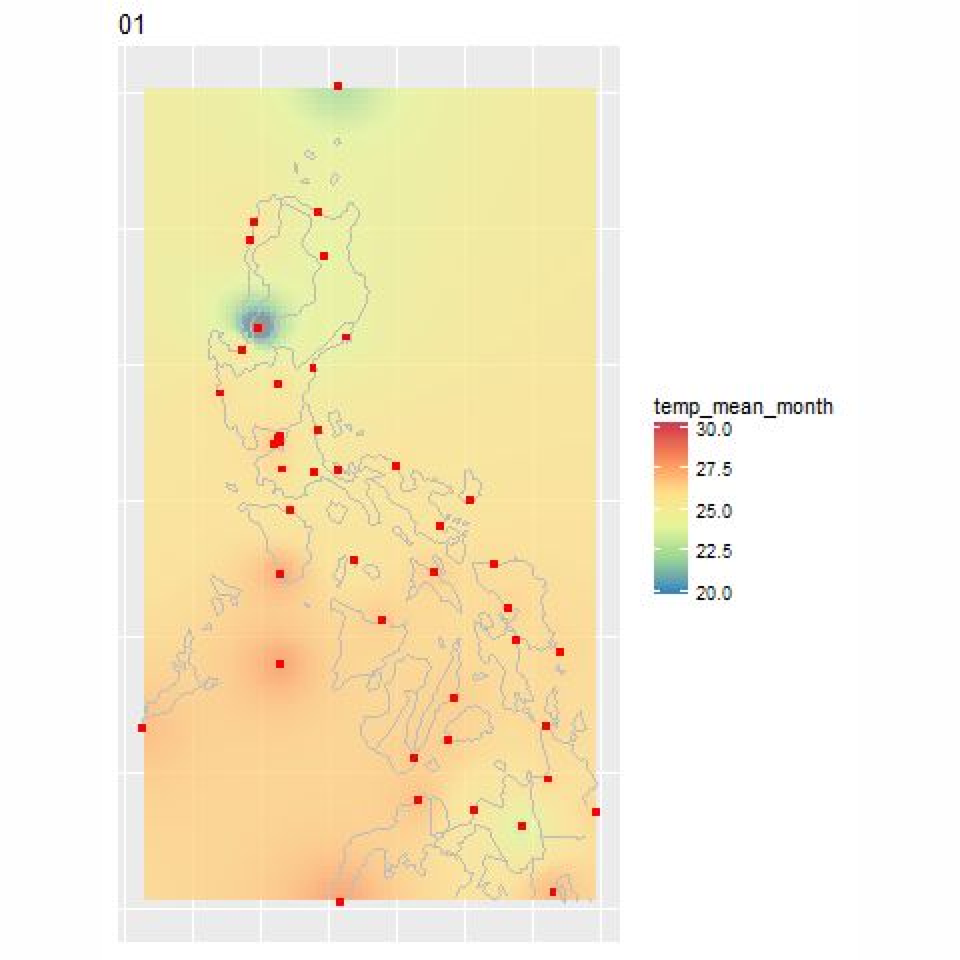
Data with missing values with replaced with mean



Data with missing values with replaced with **predictive mean matching**

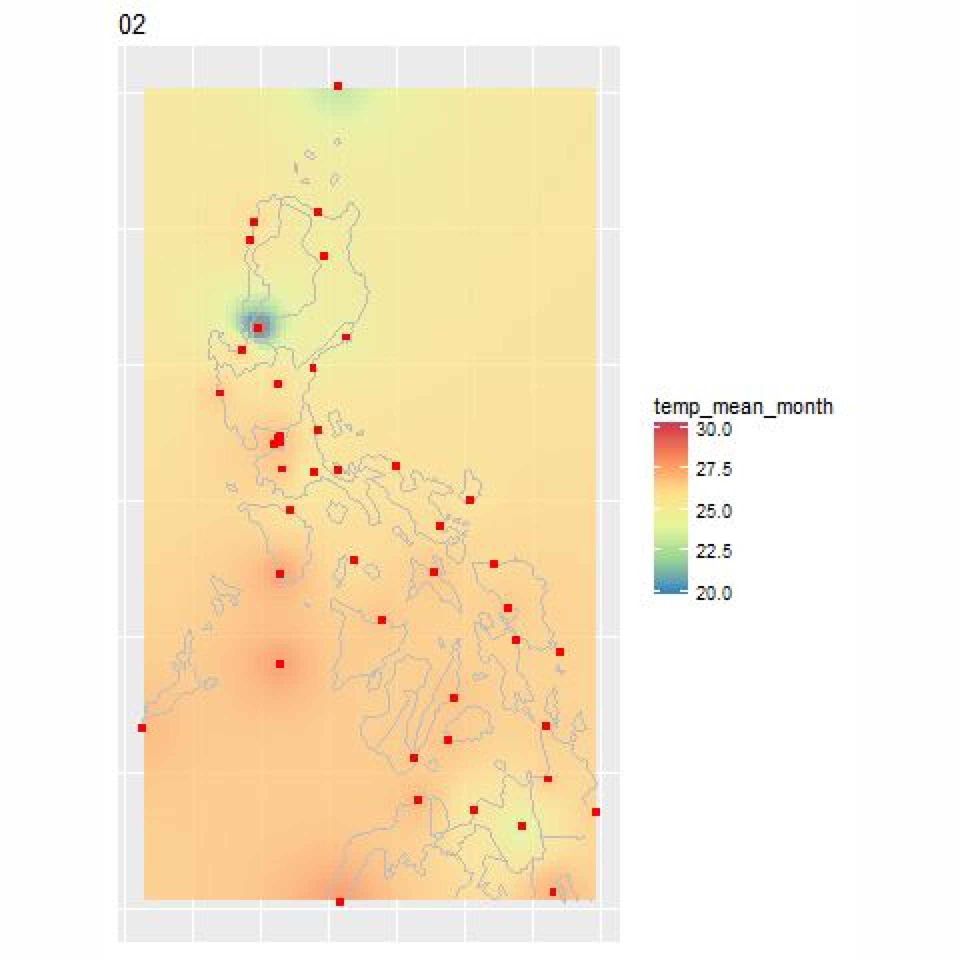
Visualizations

Temperature



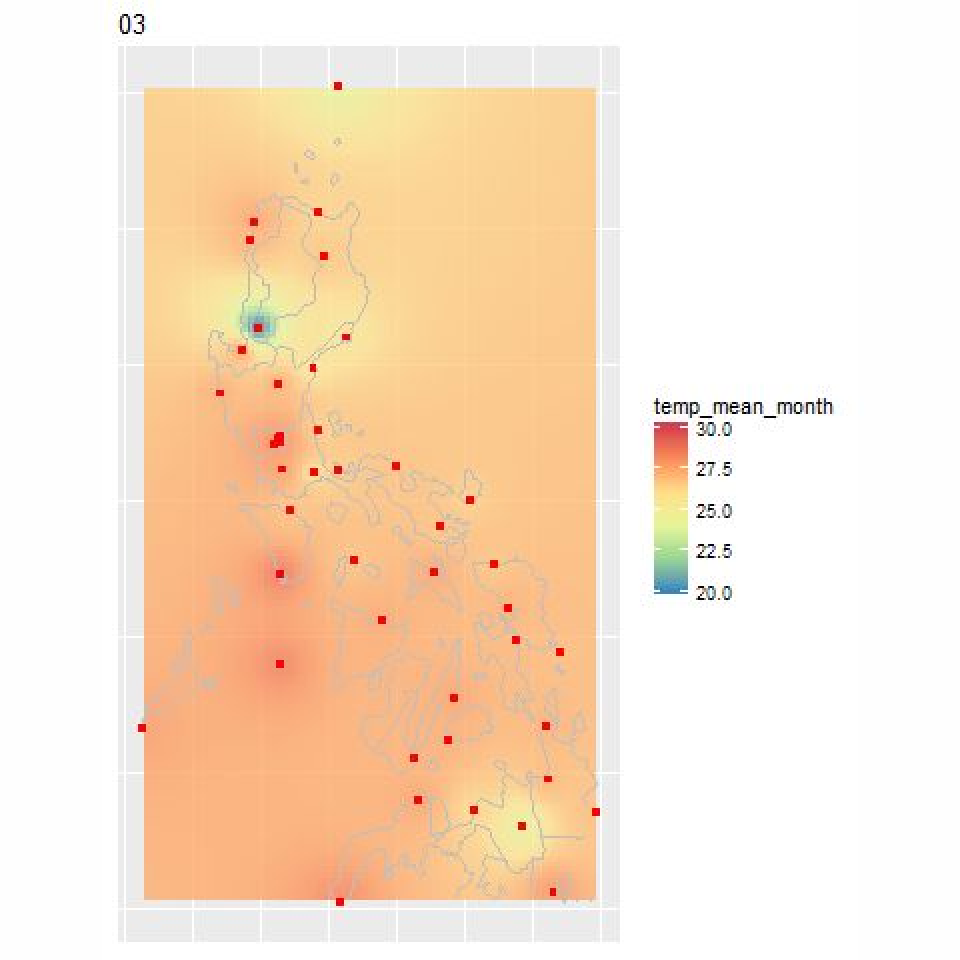
Average Monthly Temperature (°C)

January



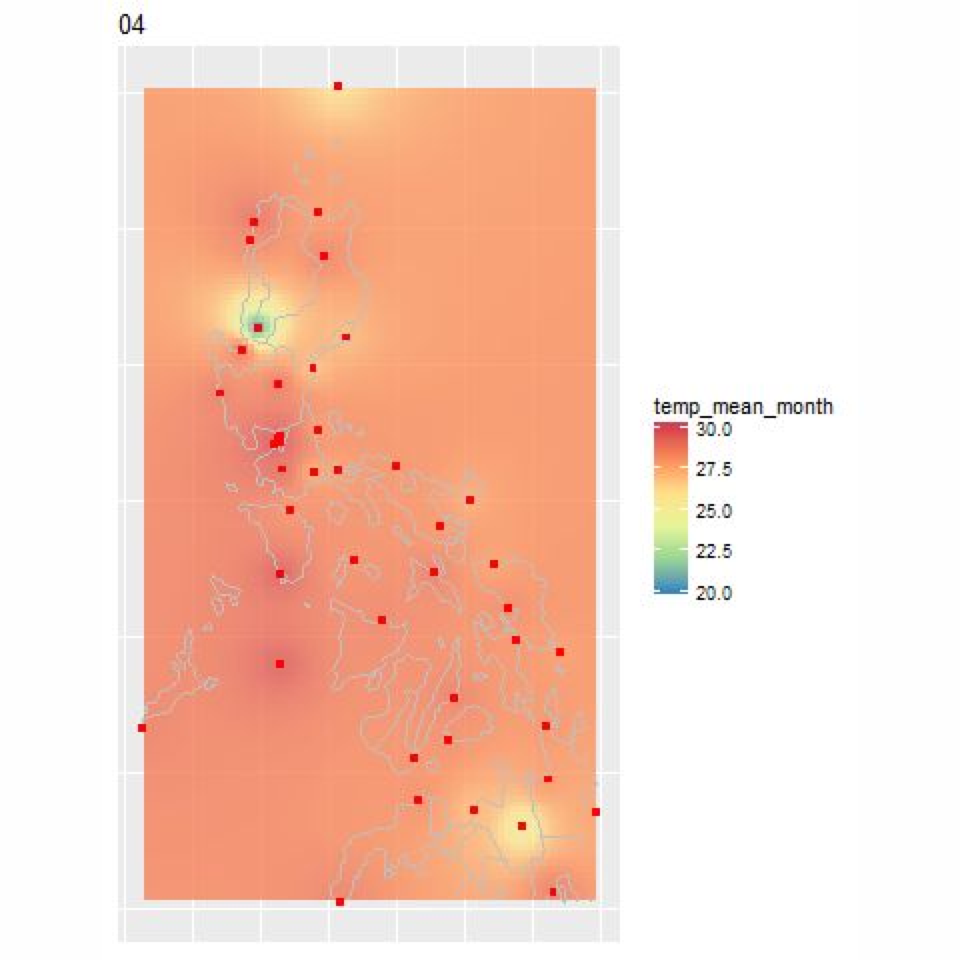
Average Monthly Temperature (°C)

February



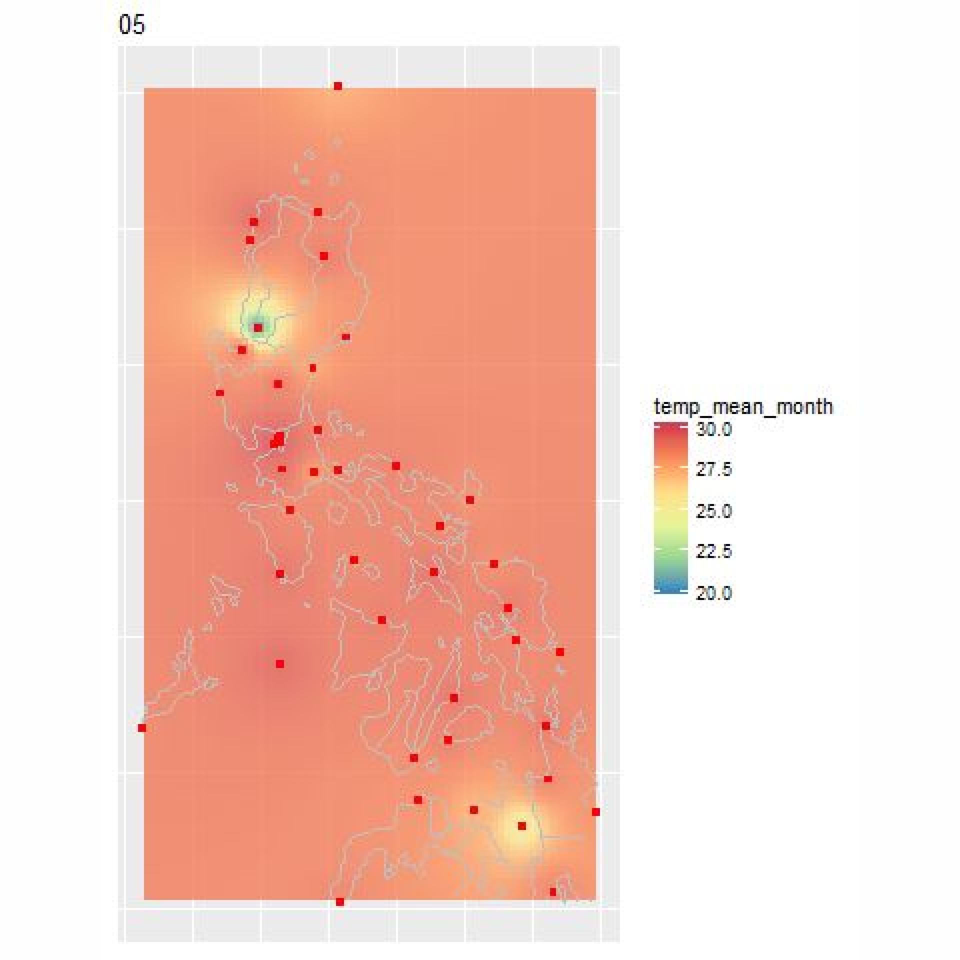
Average Monthly Temperature (°C)

March



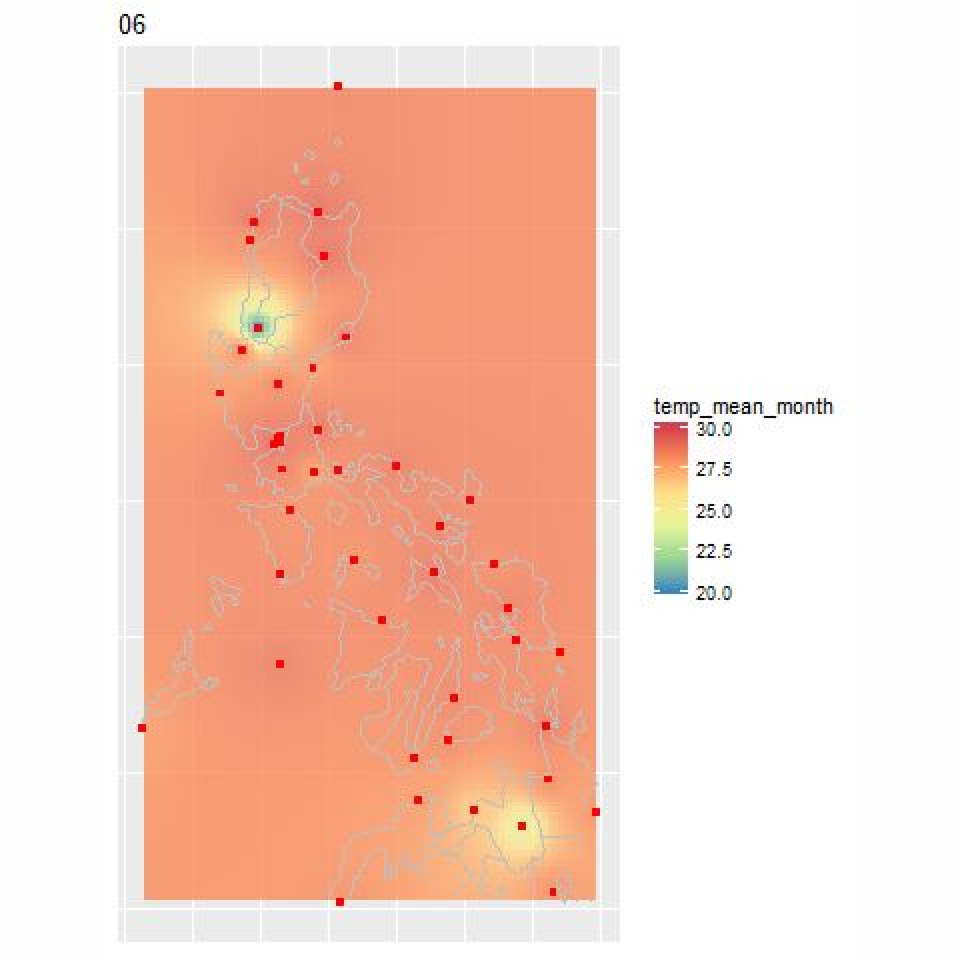
Average Monthly Temperature (°C)

April



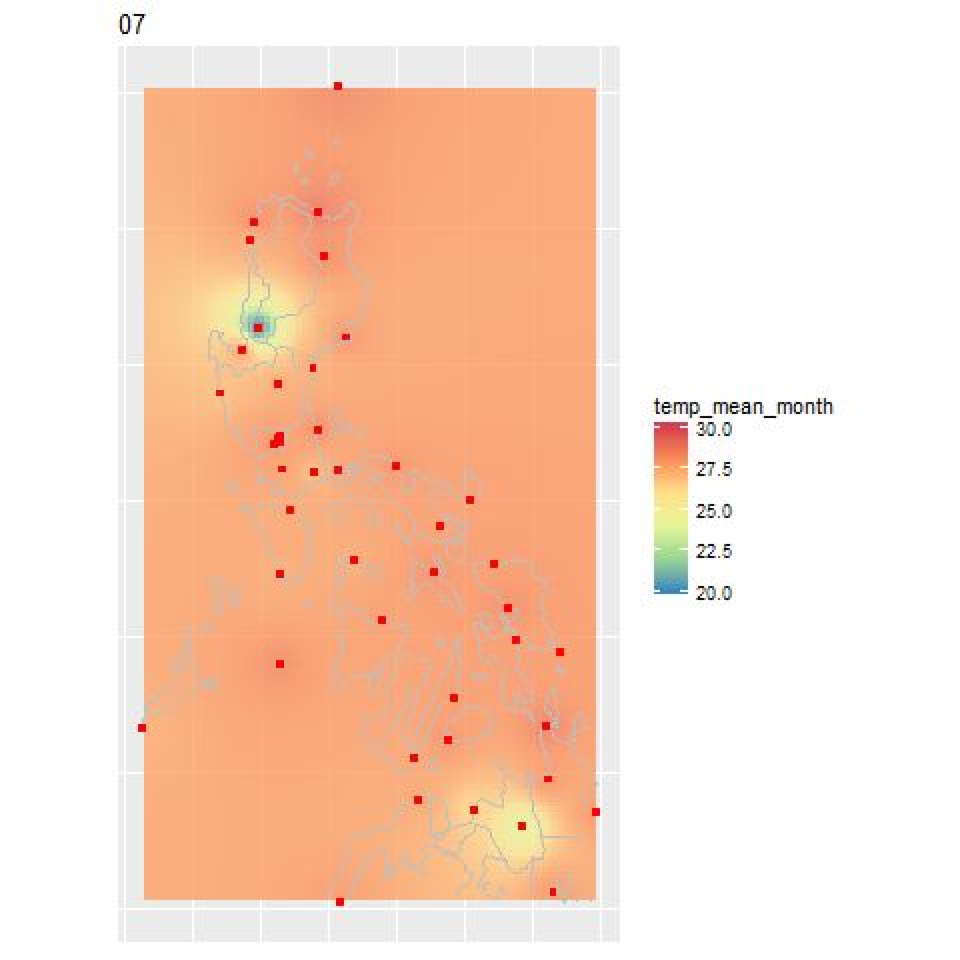
Average Monthly Temperature (°C)

May



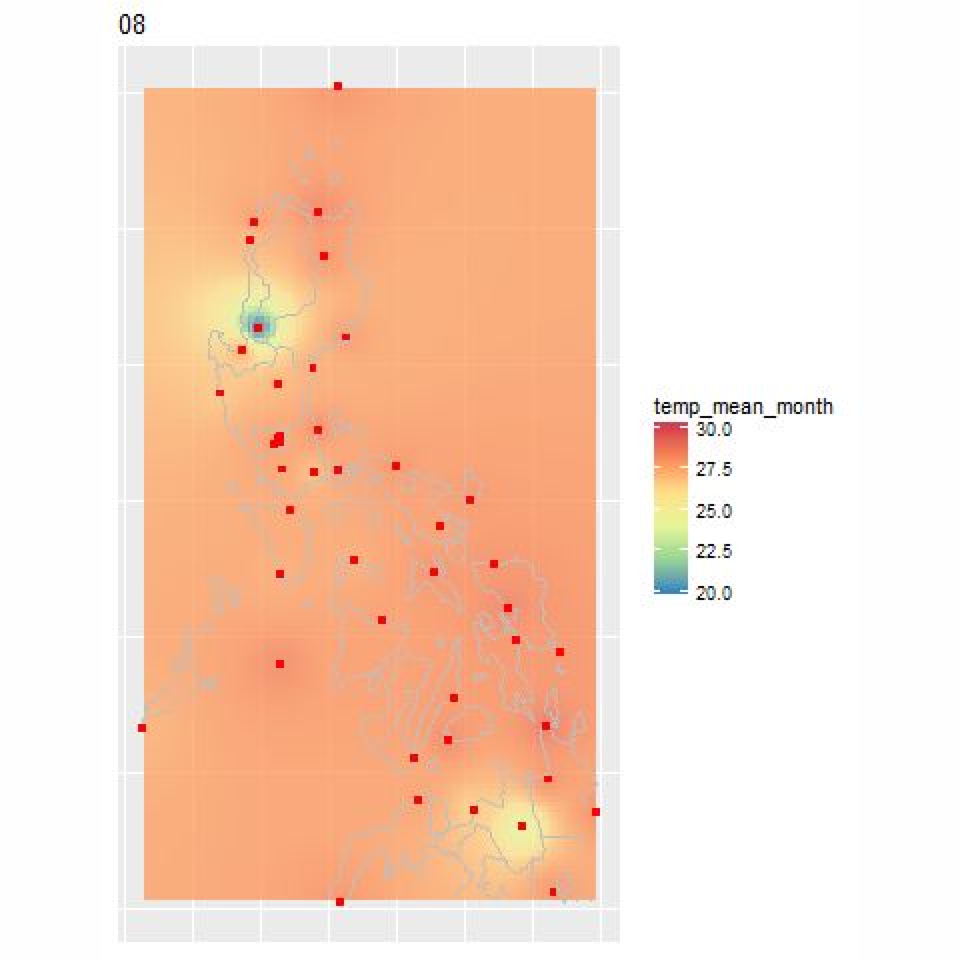
Average Monthly Temperature (°C)

June



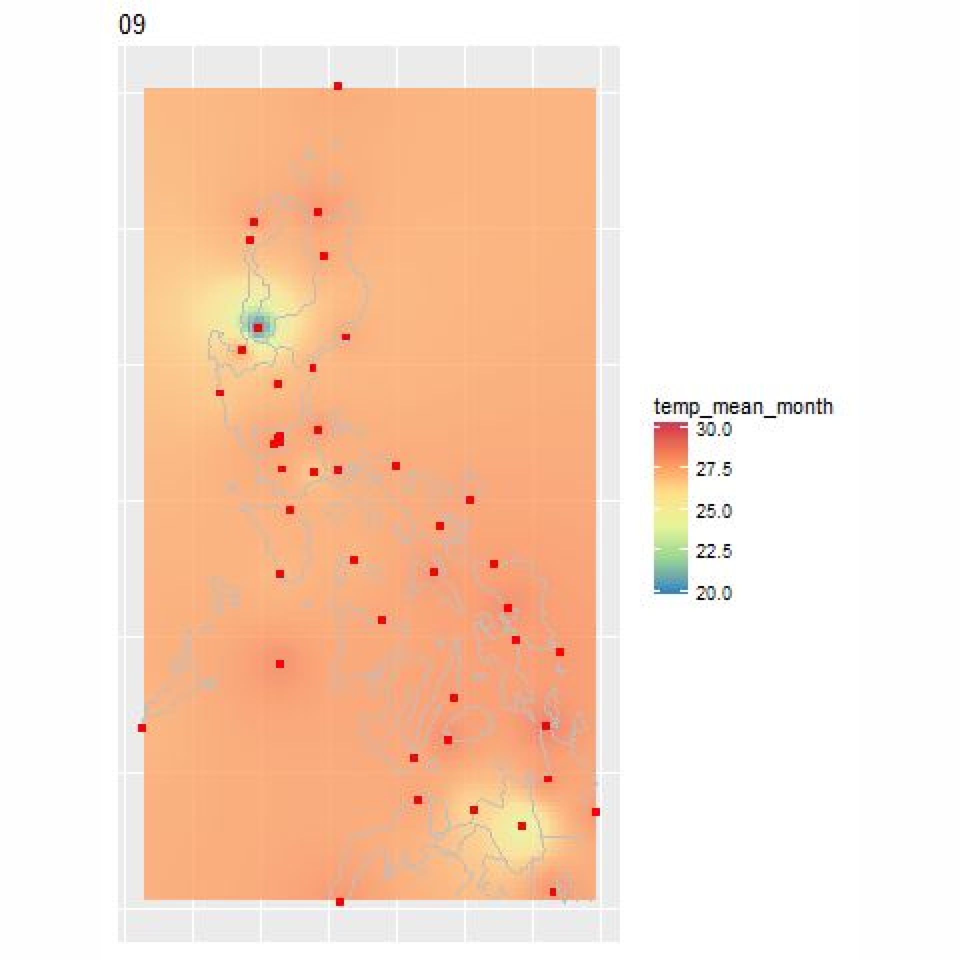
Average Monthly Temperature (°C)

July



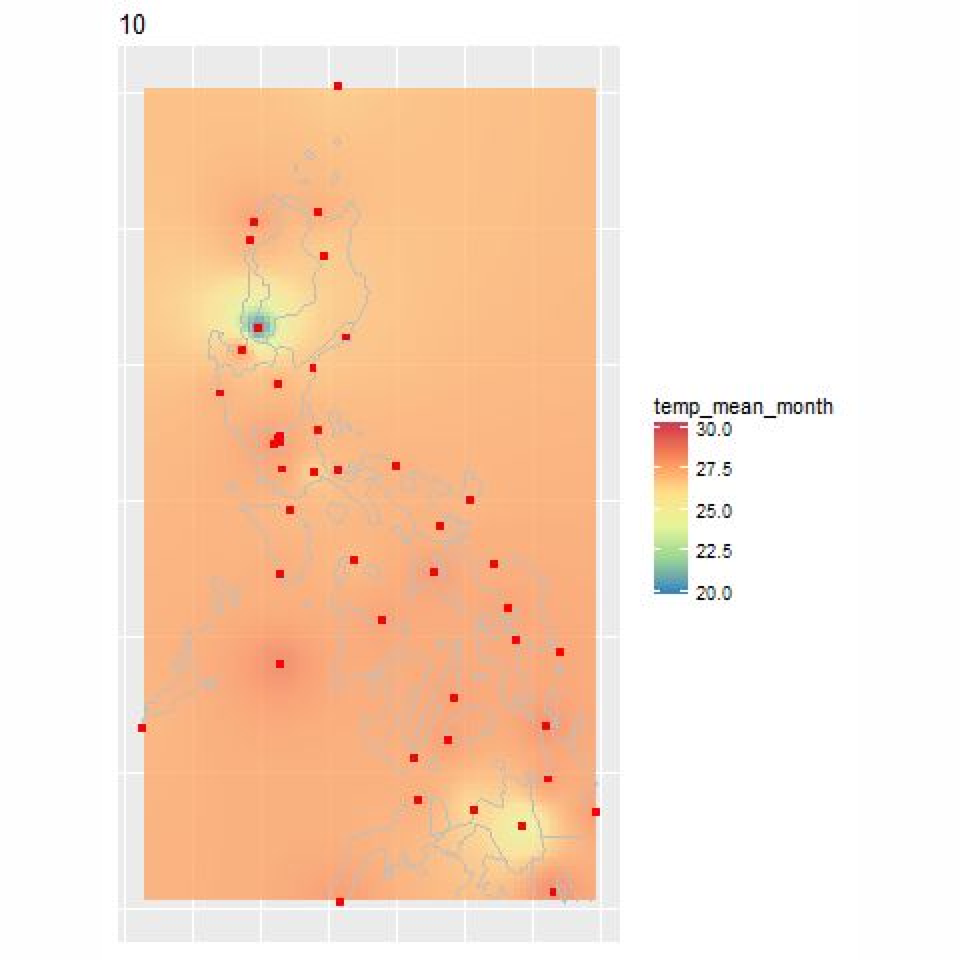
Average Monthly Temperature (°C)

August



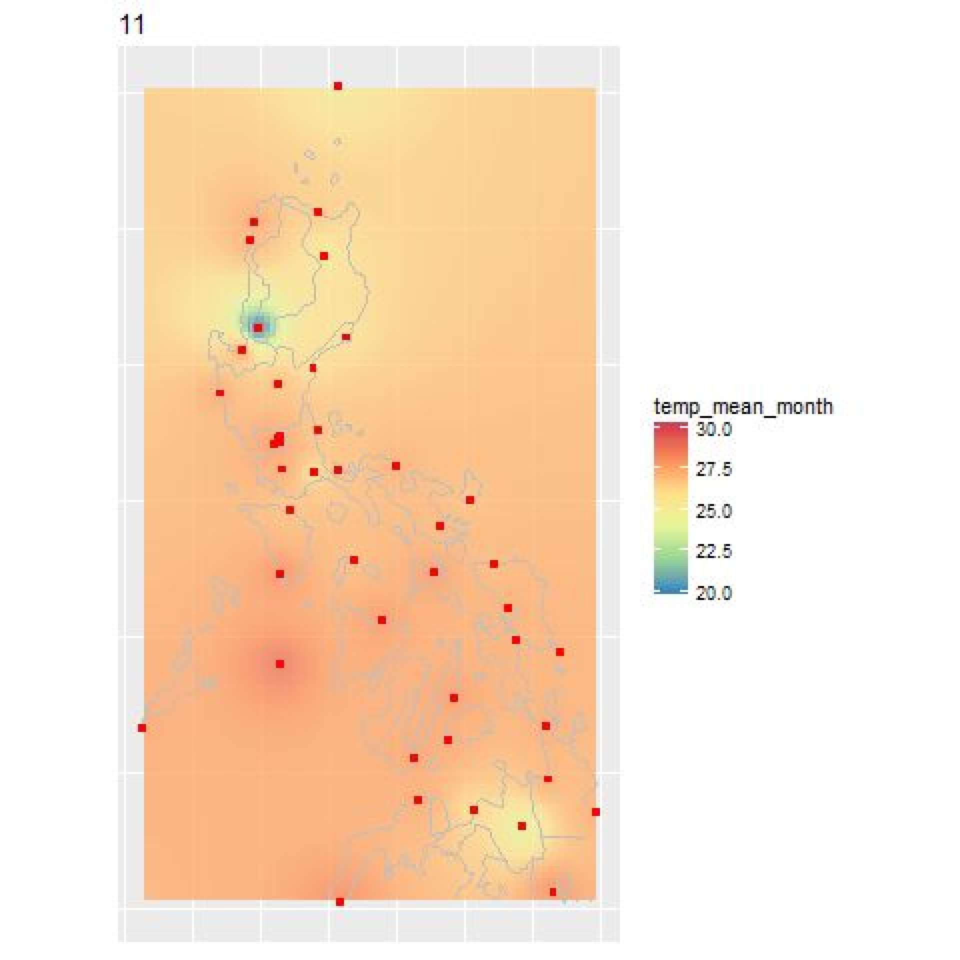
Average Monthly Temperature (°C)

September



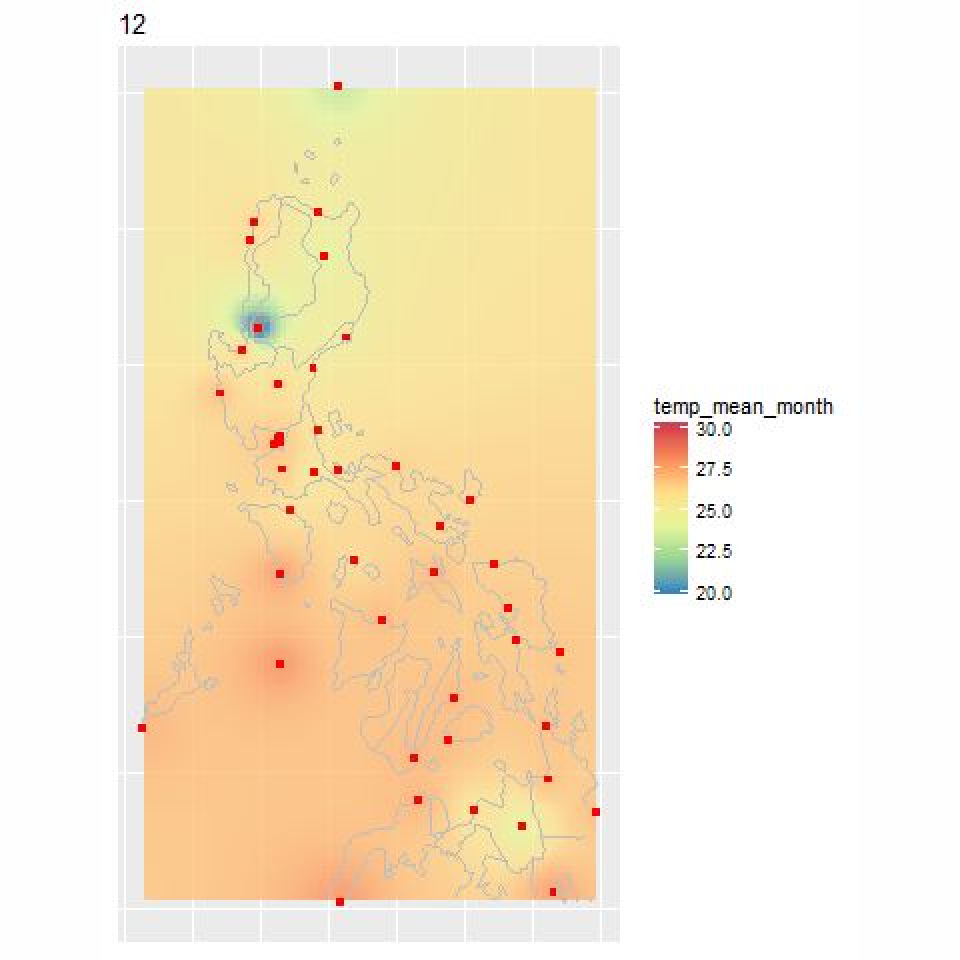
Average Monthly Temperature (°C)

October



Average Monthly Temperature (°C)

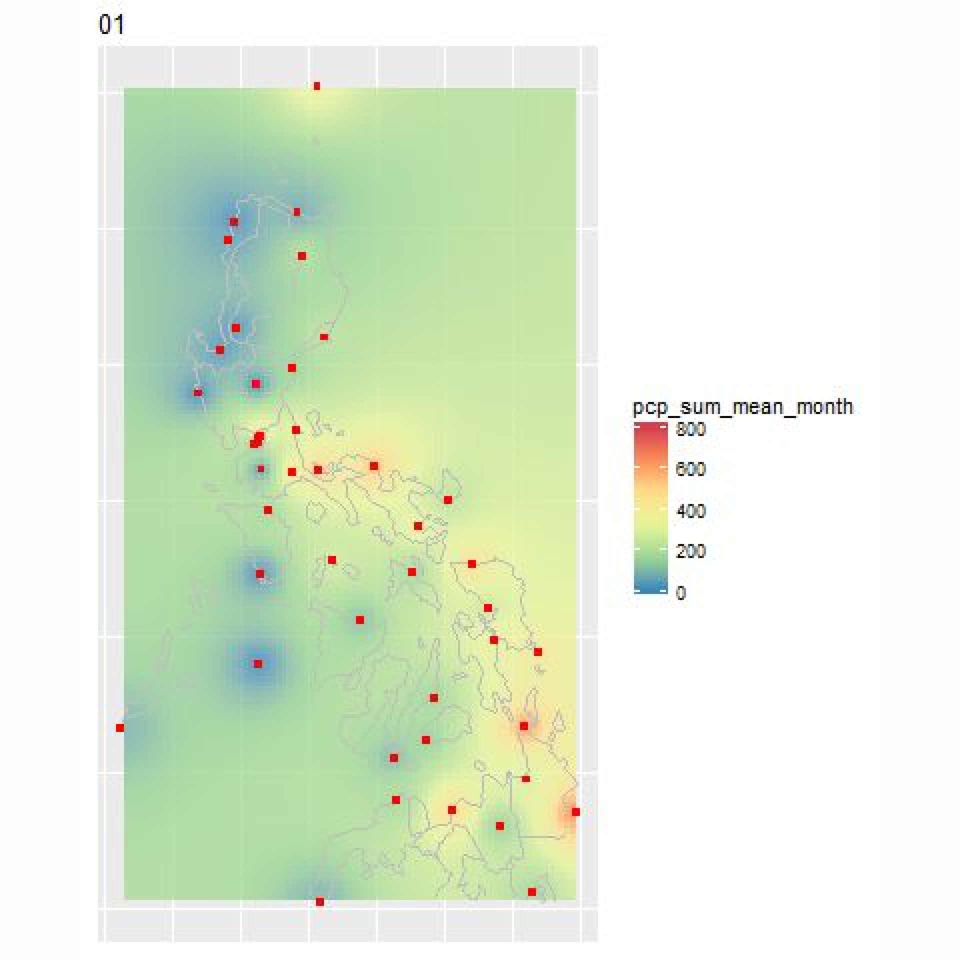
November



Average Monthly Temperature (°C)

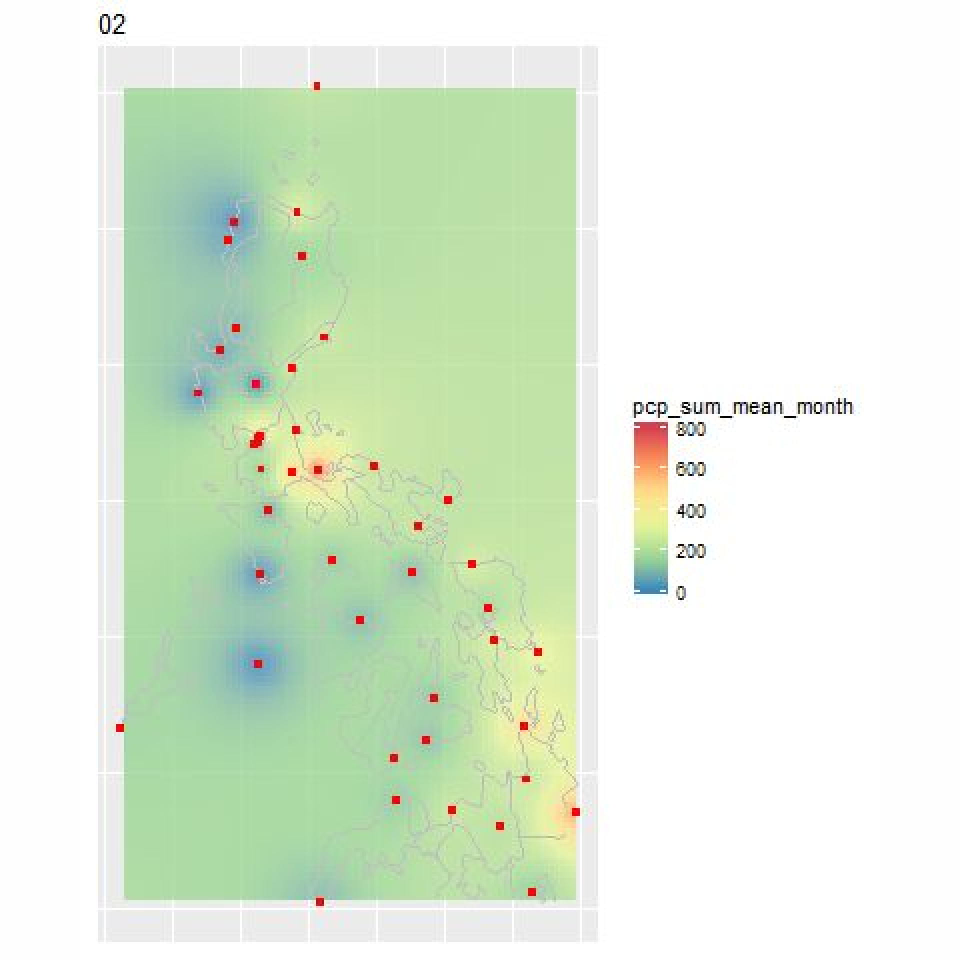
December

Precipitation



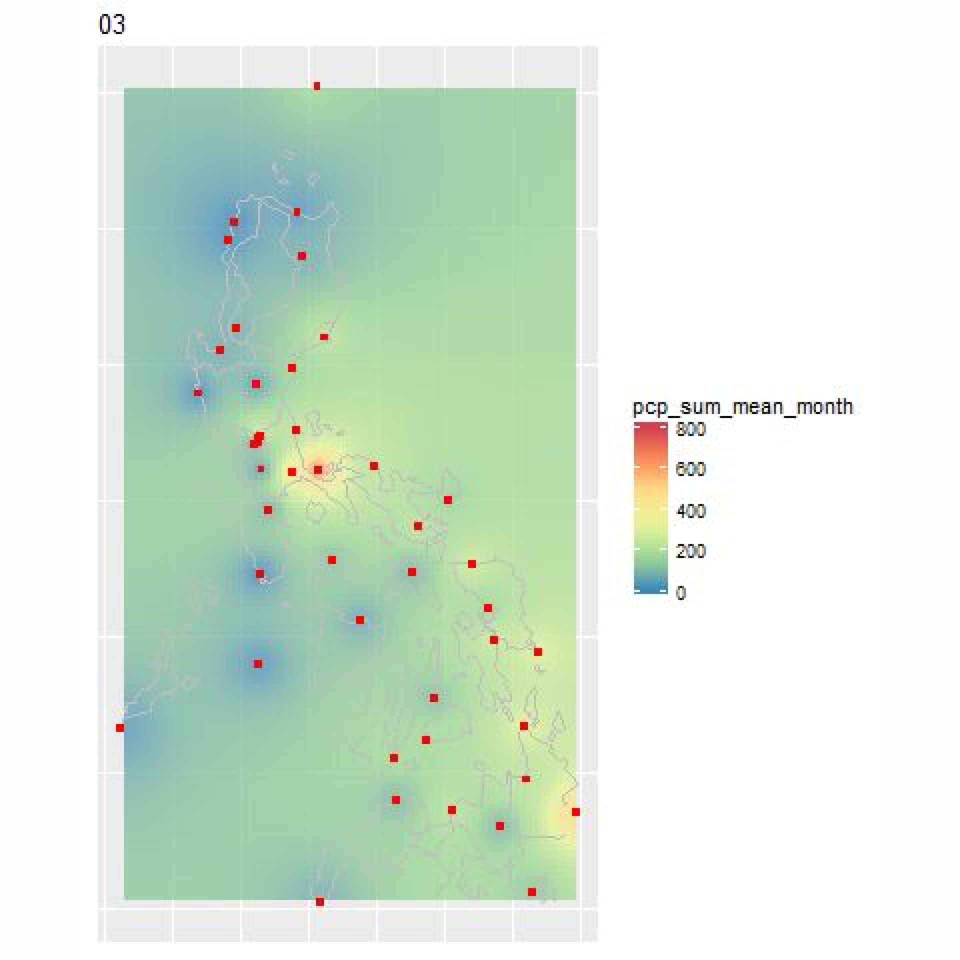
Average Monthly Precipitation (mm)

January



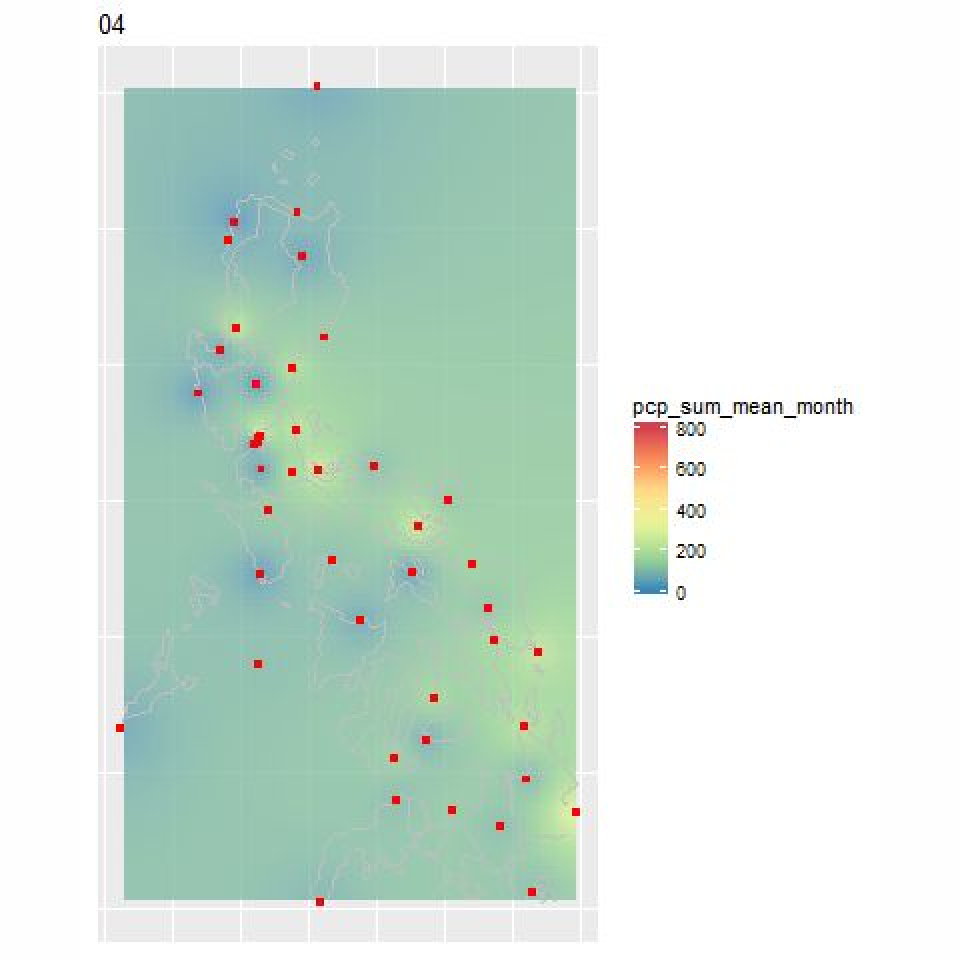
Average Monthly Precipitation (mm)

February



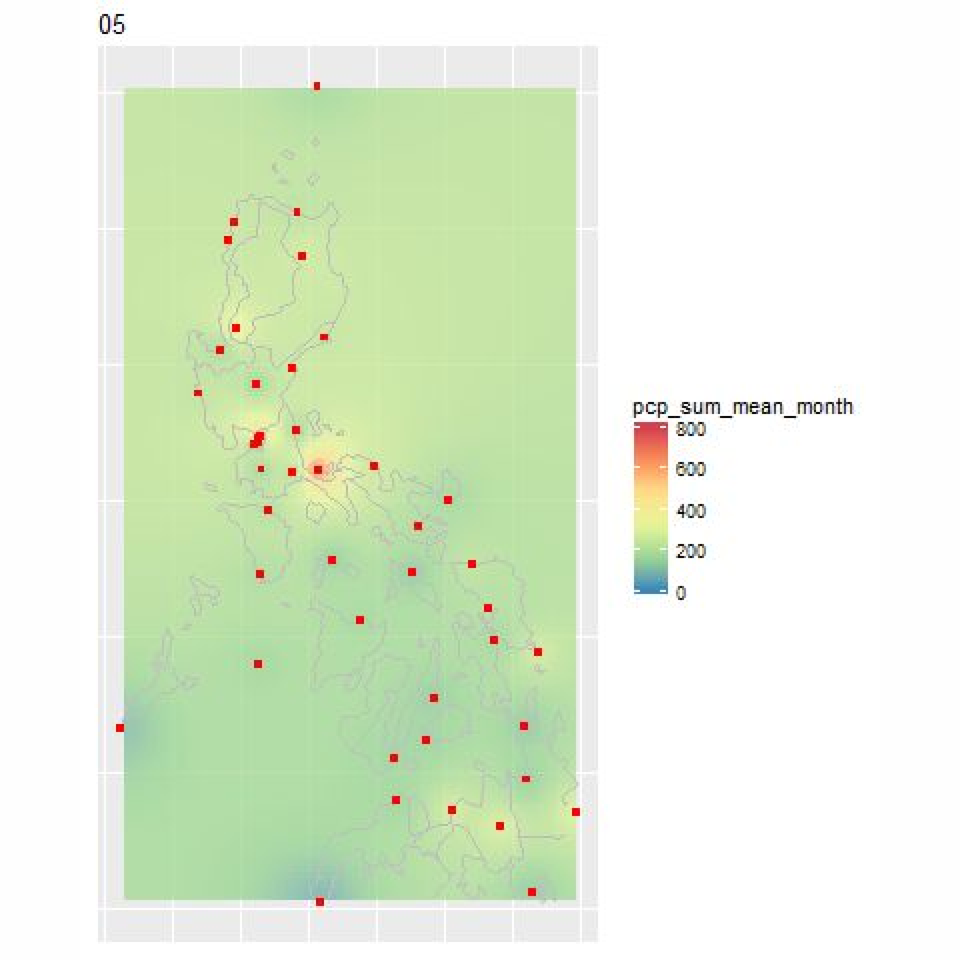
Average Monthly Precipitation (mm)

March



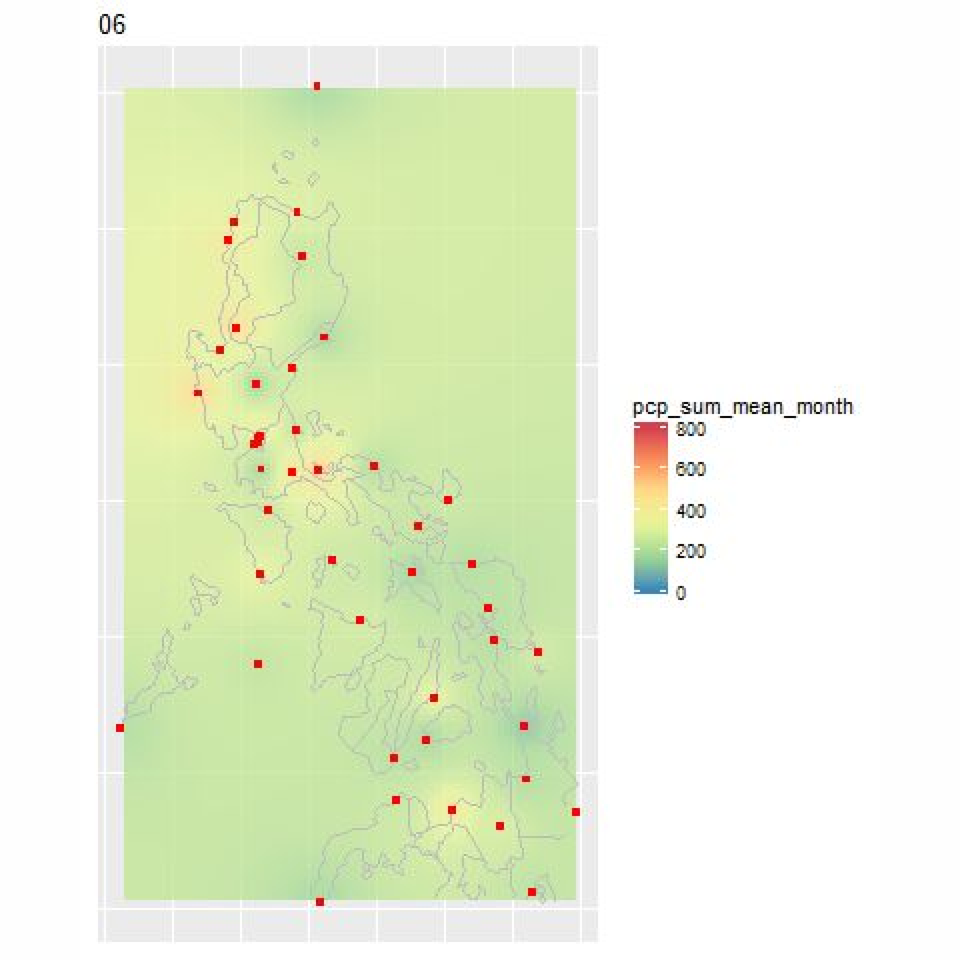
Average Monthly Precipitation (mm)

April



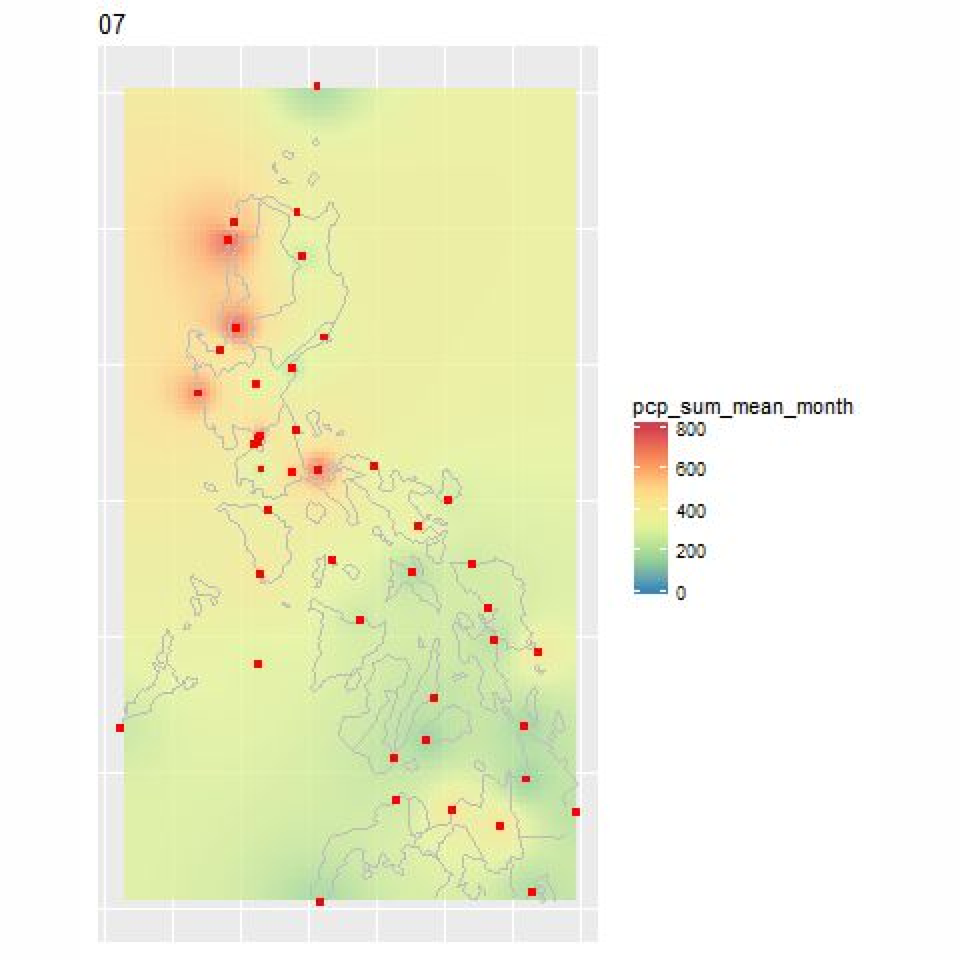
Average Monthly Precipitation (mm)

May



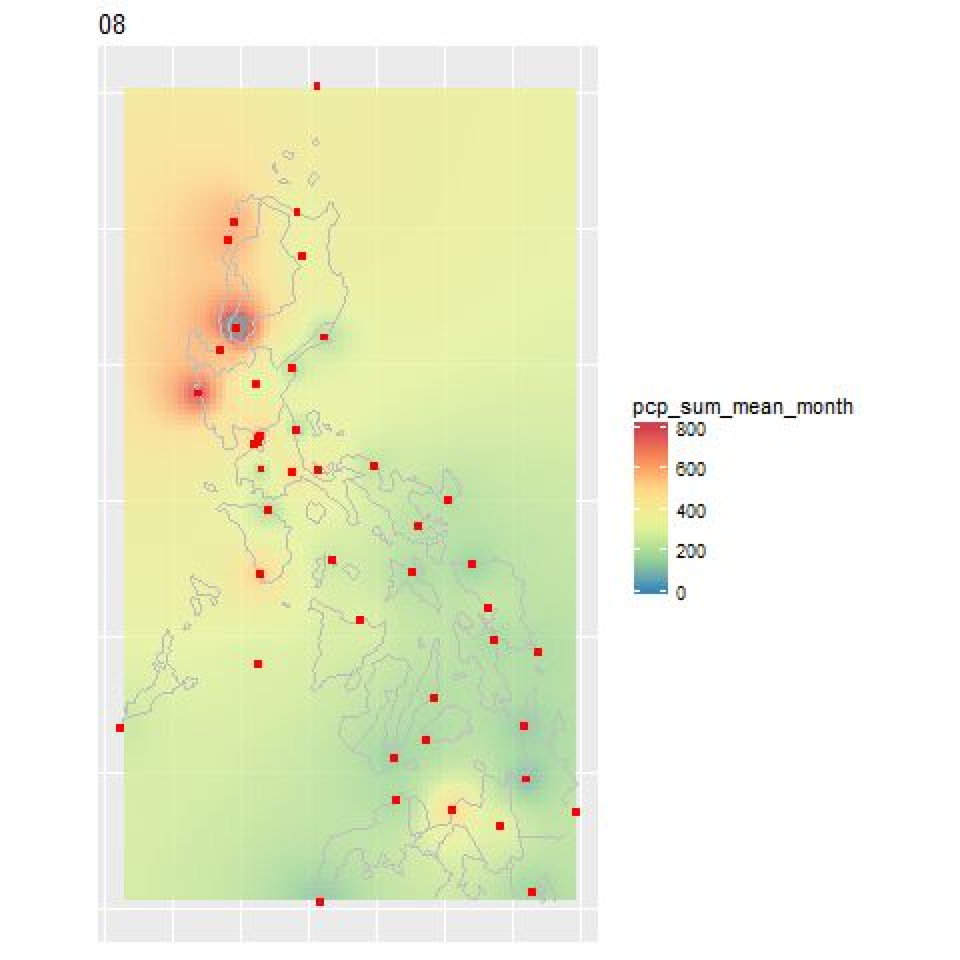
Average Monthly Precipitation (mm)

June



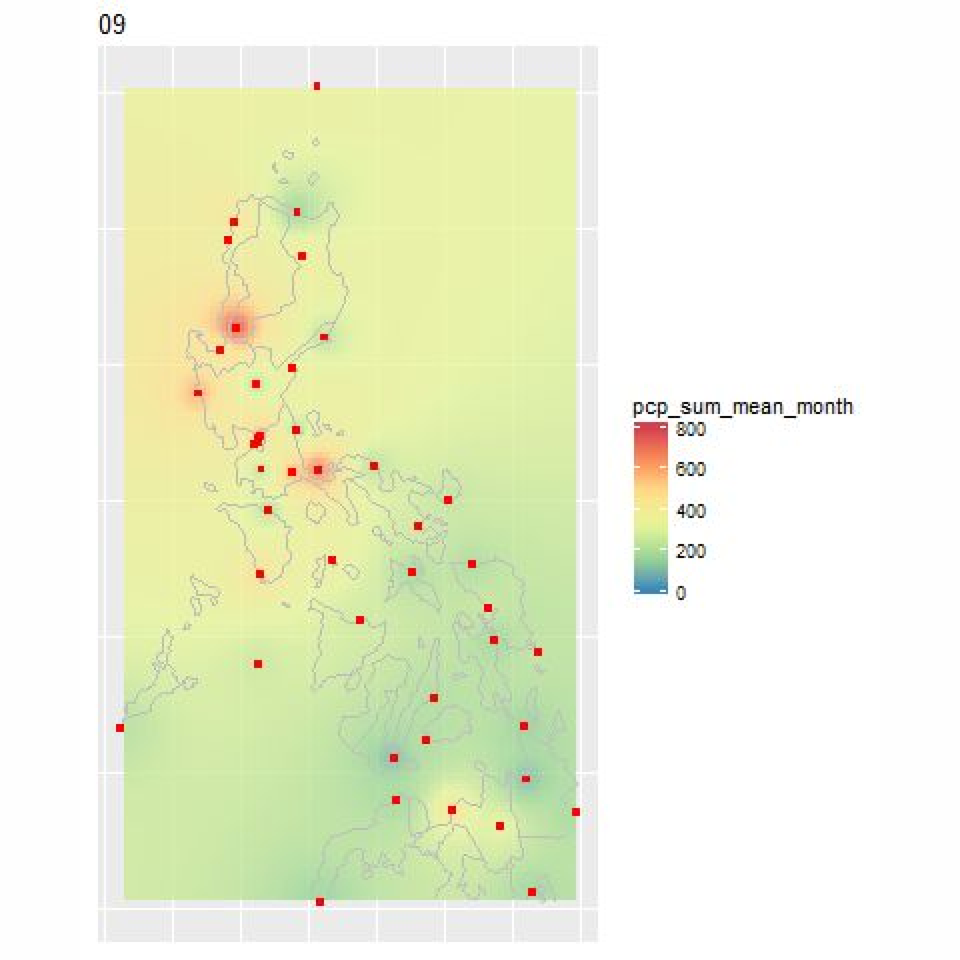
Average Monthly Precipitation (mm)

July



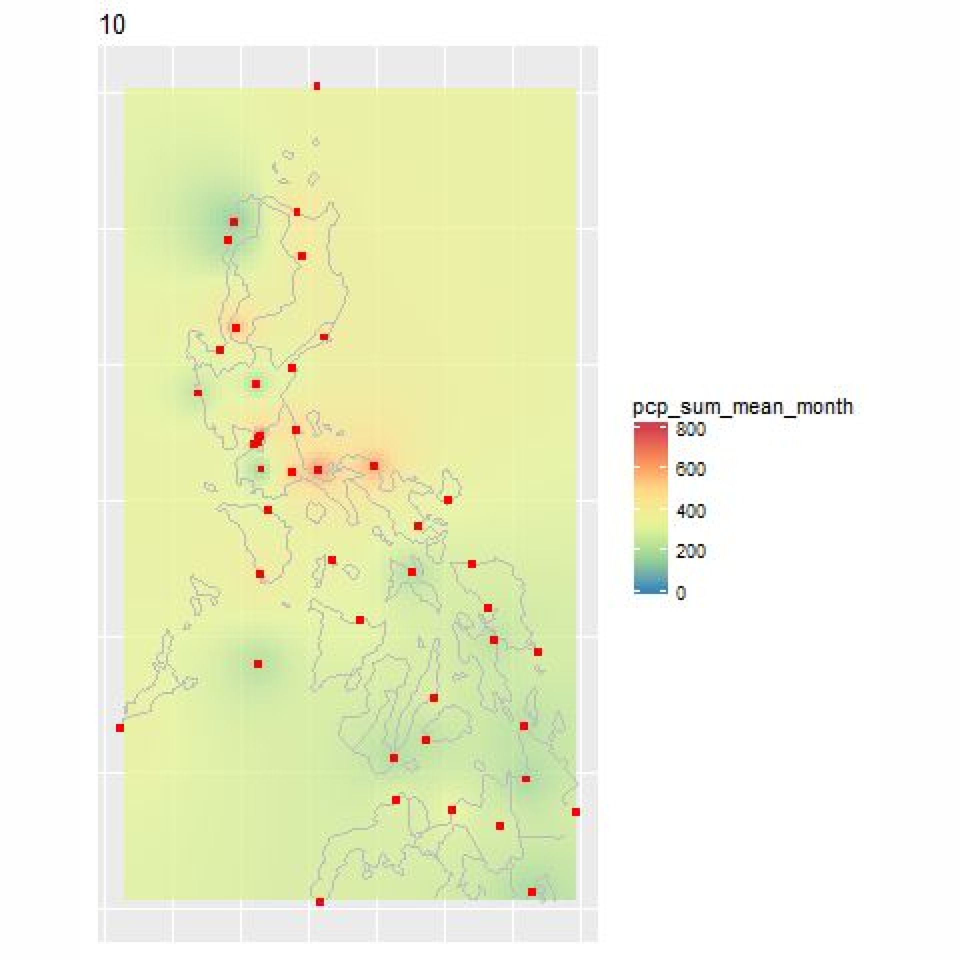
Average Monthly Precipitation (mm)

August



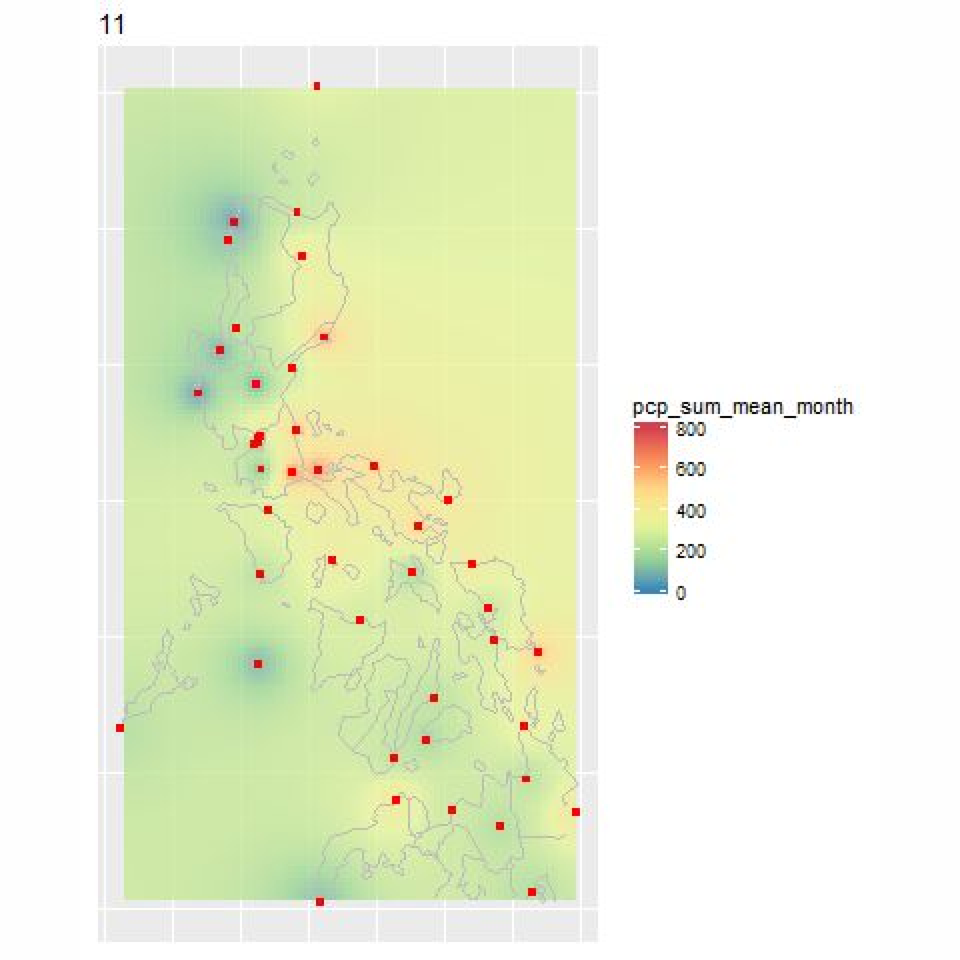
Average Monthly Precipitation (mm)

September



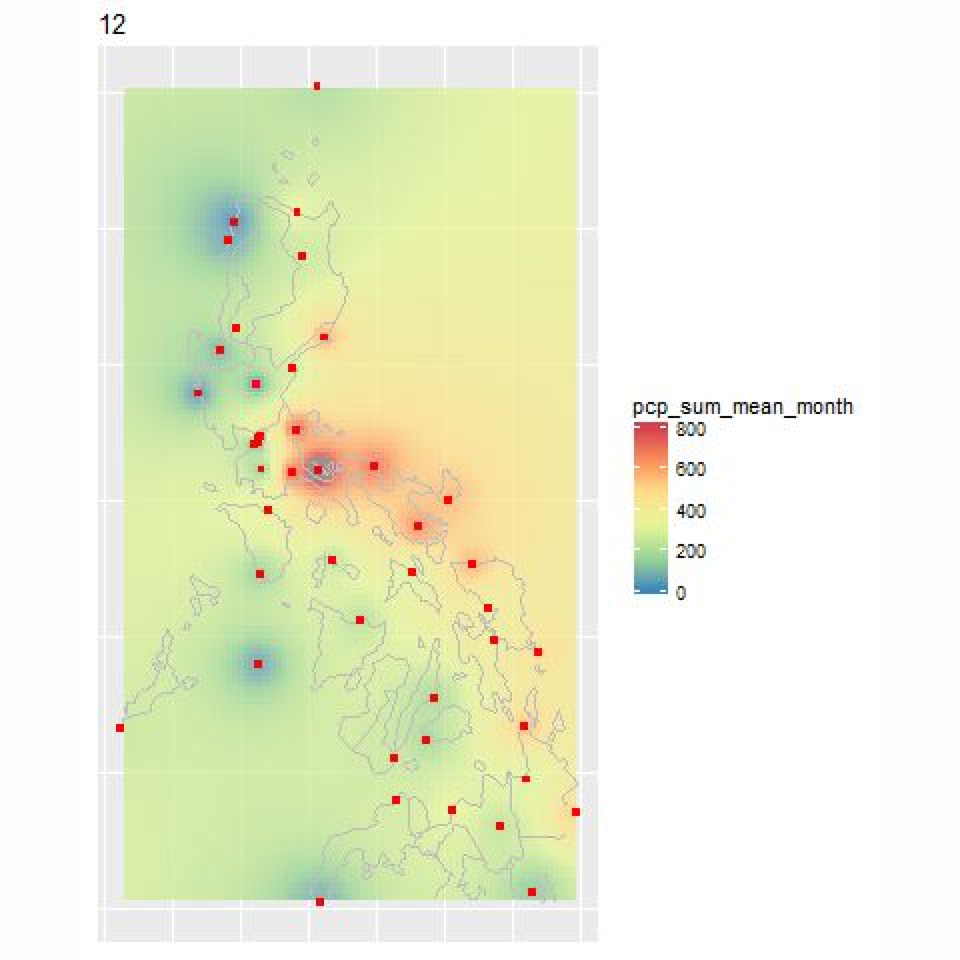
Average Monthly Precipitation (mm)

October



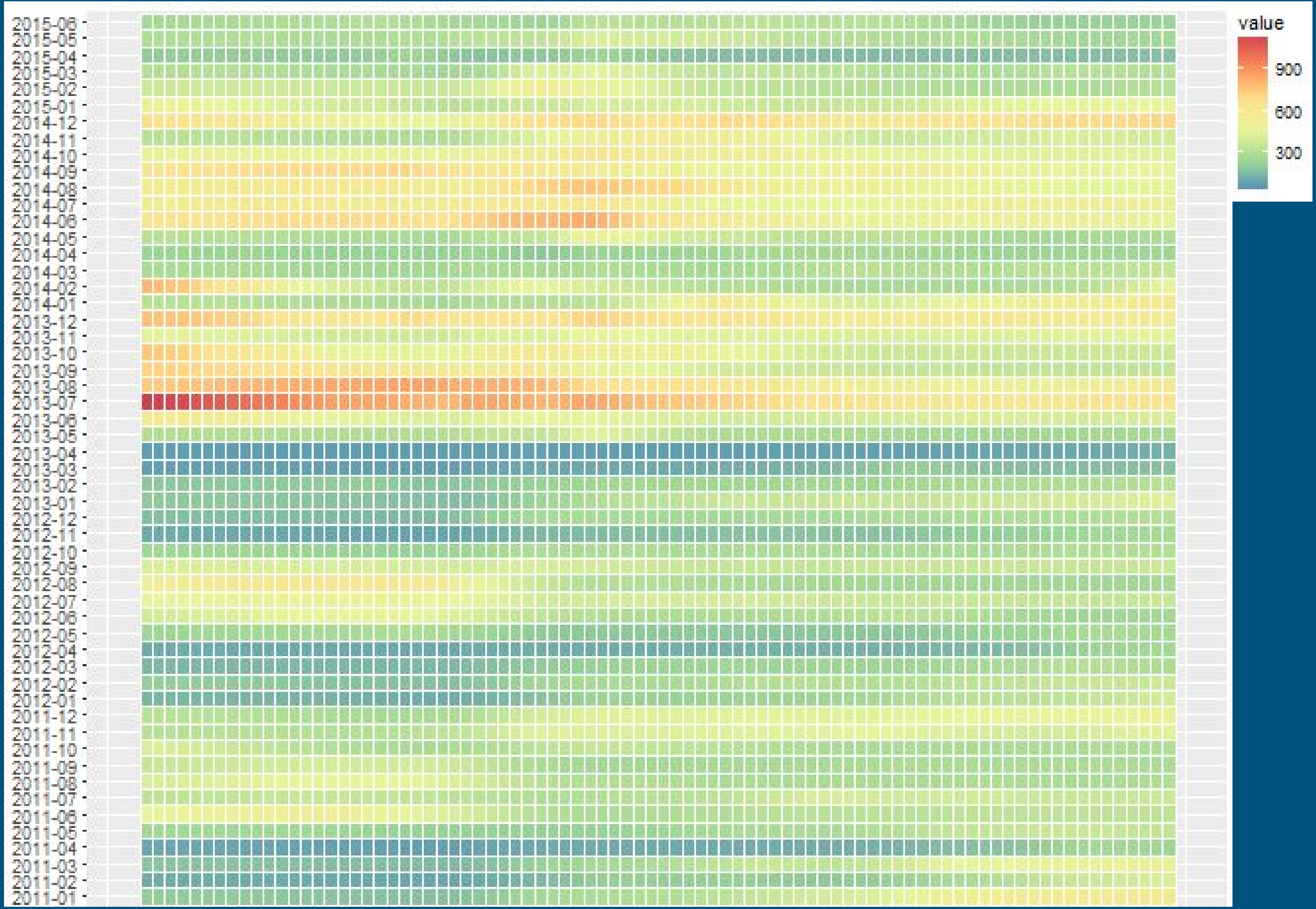
Average Monthly Precipitation (mm)

November



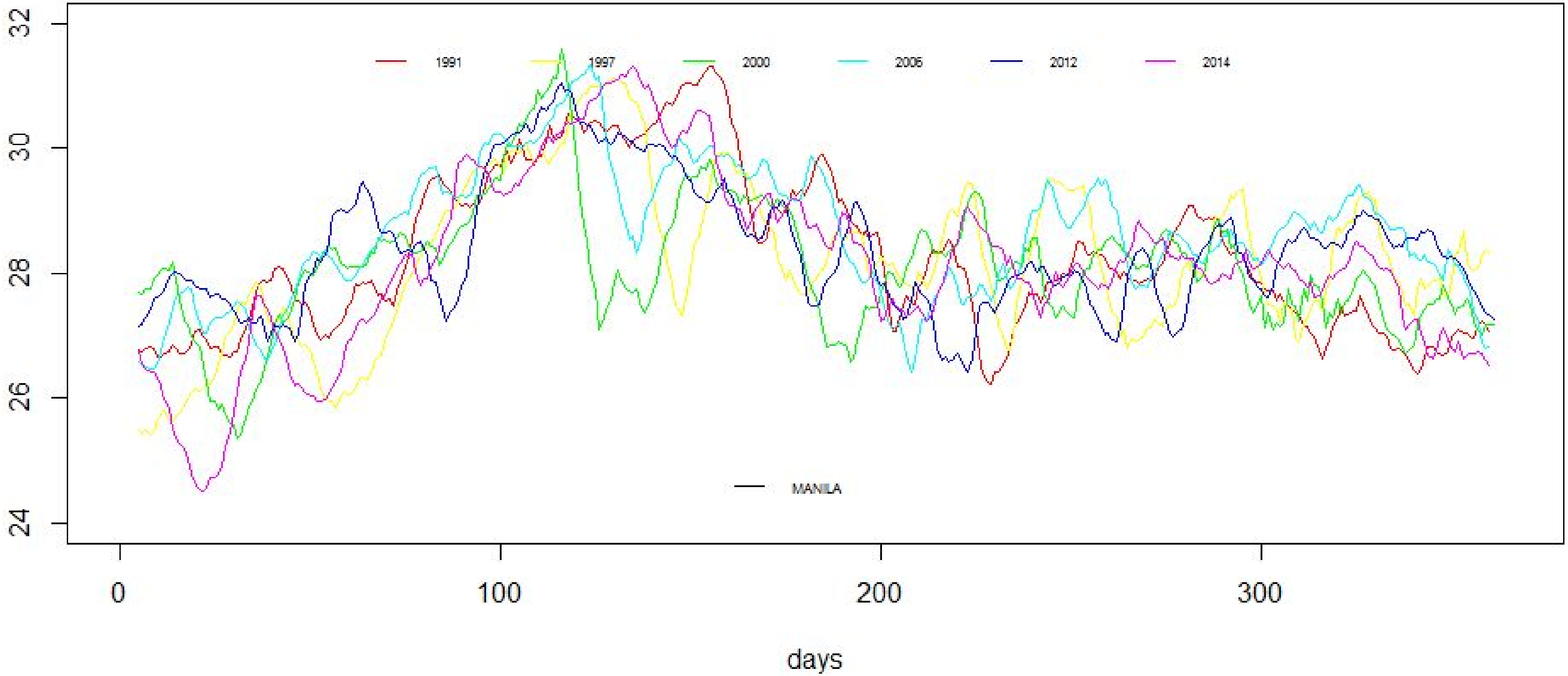
Average Monthly Precipitation (mm)

December

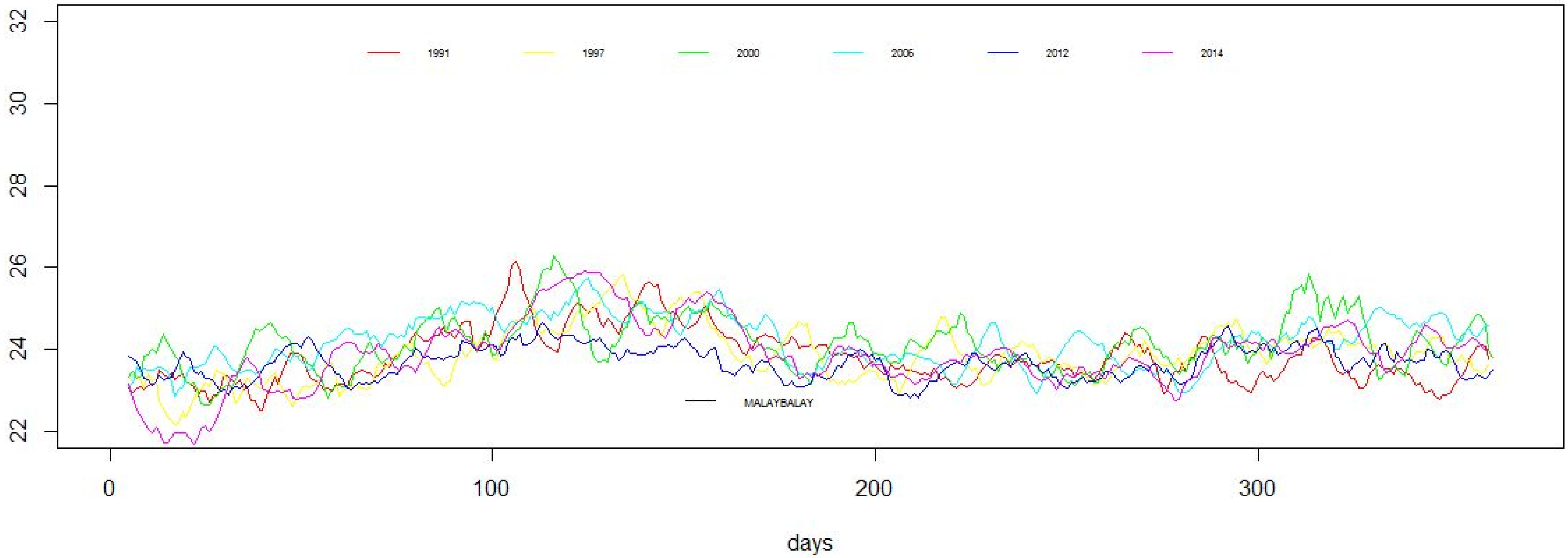

2015-06
2015-05
2015-04
2015-03
2015-02
2015-01
2014-12
2014-11
2014-10
2014-09
2014-08
2014-07
2014-06
2014-05
2014-04
2014-03
2014-02
2014-01
2013-12
2013-11
2013-10
2013-09
2013-08
2013-07
2013-06
2013-05
2013-04
2013-03
2013-02
2013-01
2012-12
2012-11
2012-10
2012-09
2012-08
2012-07
2012-06
2012-05
2012-04
2012-03
2012-02
2012-01
2011-12
2011-11
2011-10
2011-09
2011-08
2011-07
2011-06
2011-05
2011-04
2011-03
2011-02
2011-01
value
900
600
300

Temperature Time Series

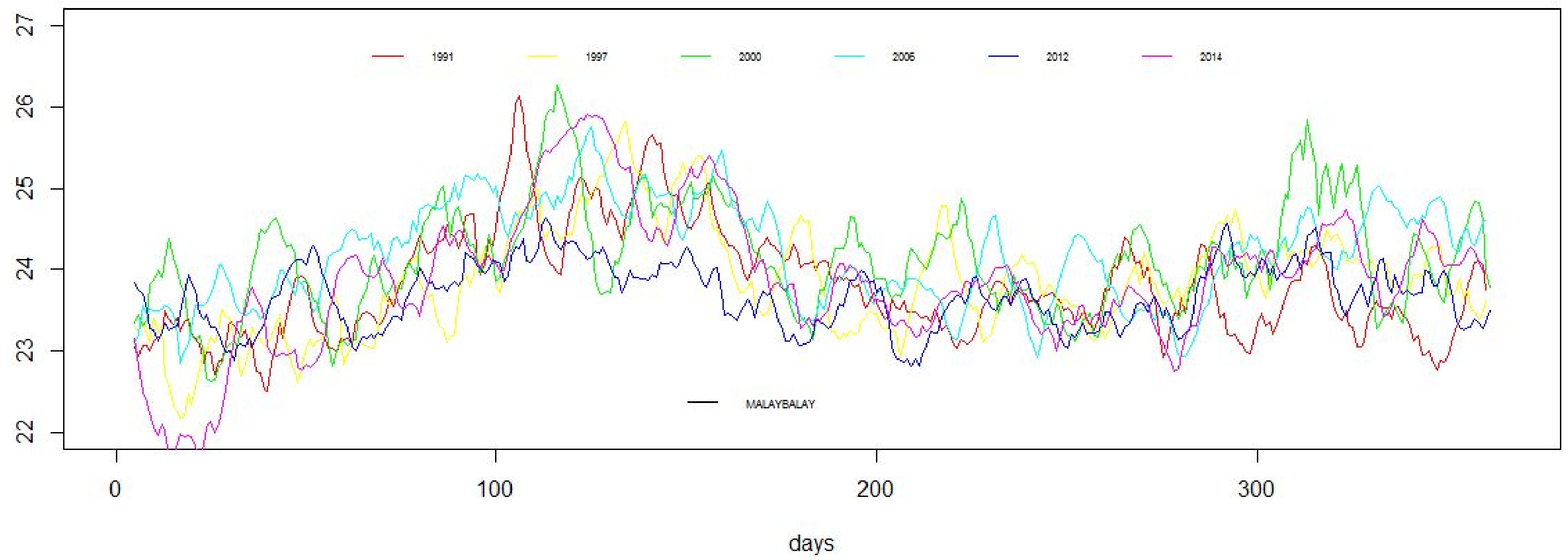
MANILA Temperature

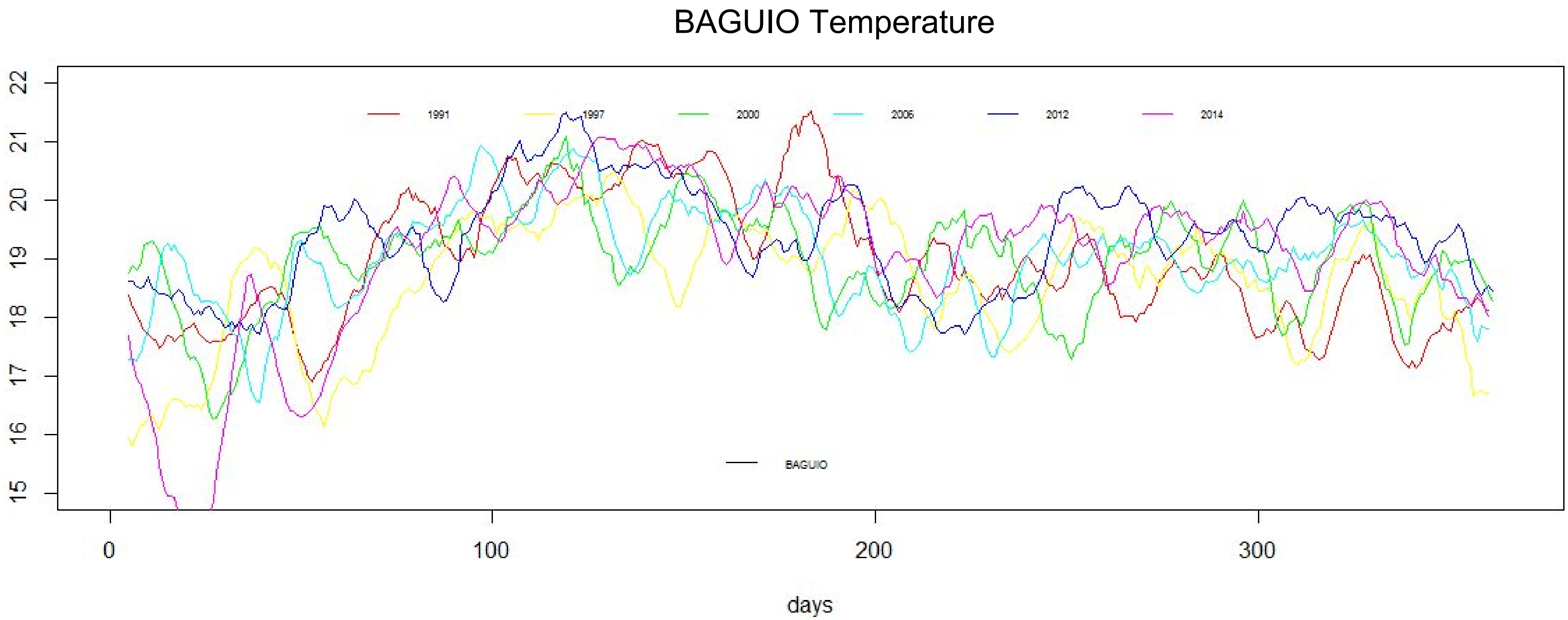


MALAYBALAY Temperature

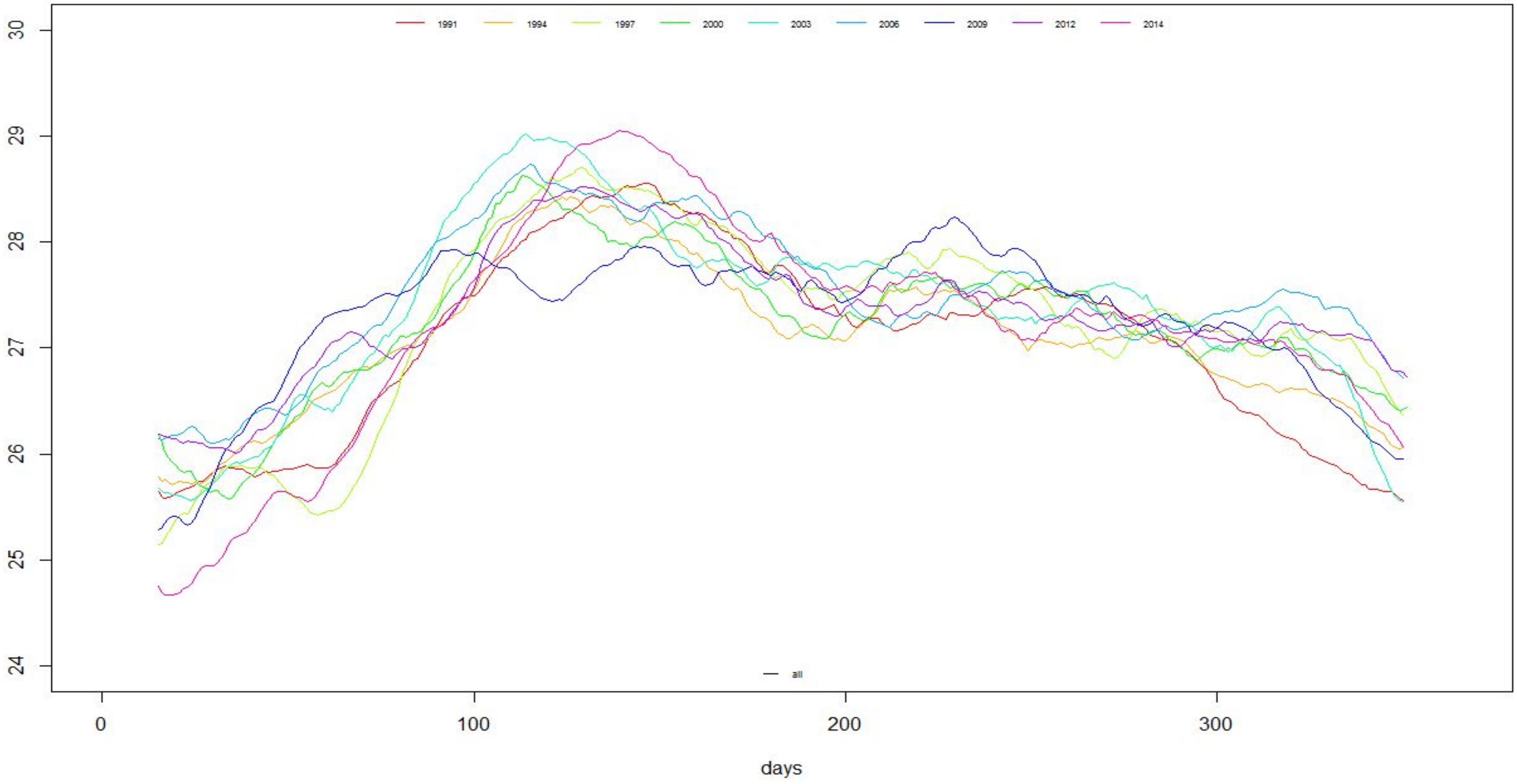


MALAYBALAY Temperature

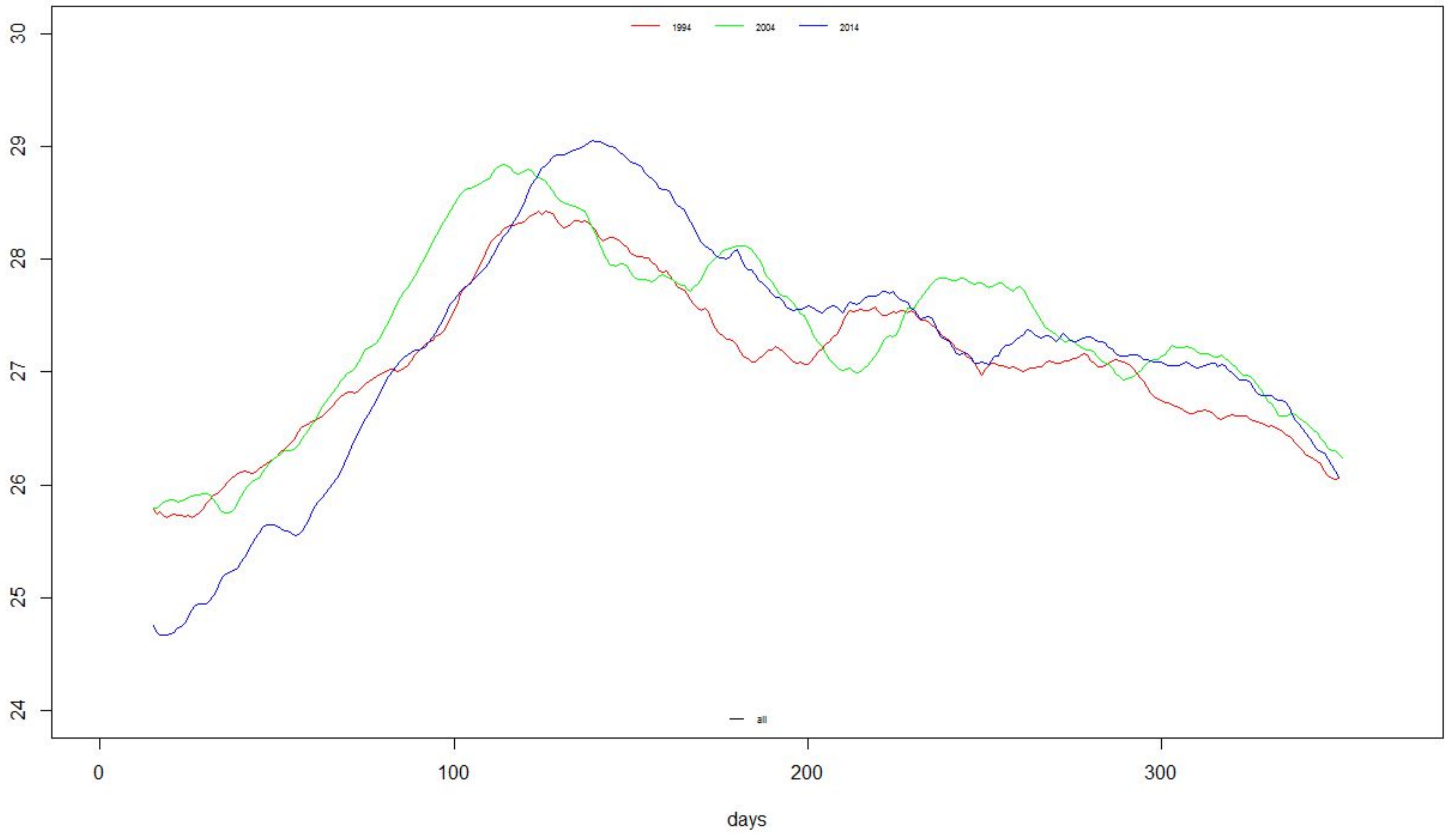



BAGUIO Temperature
1991
1997
2000
2006
2012
2014
BAGUIO
22
21
20
19
18
17
16
15
0
100
200
300
days

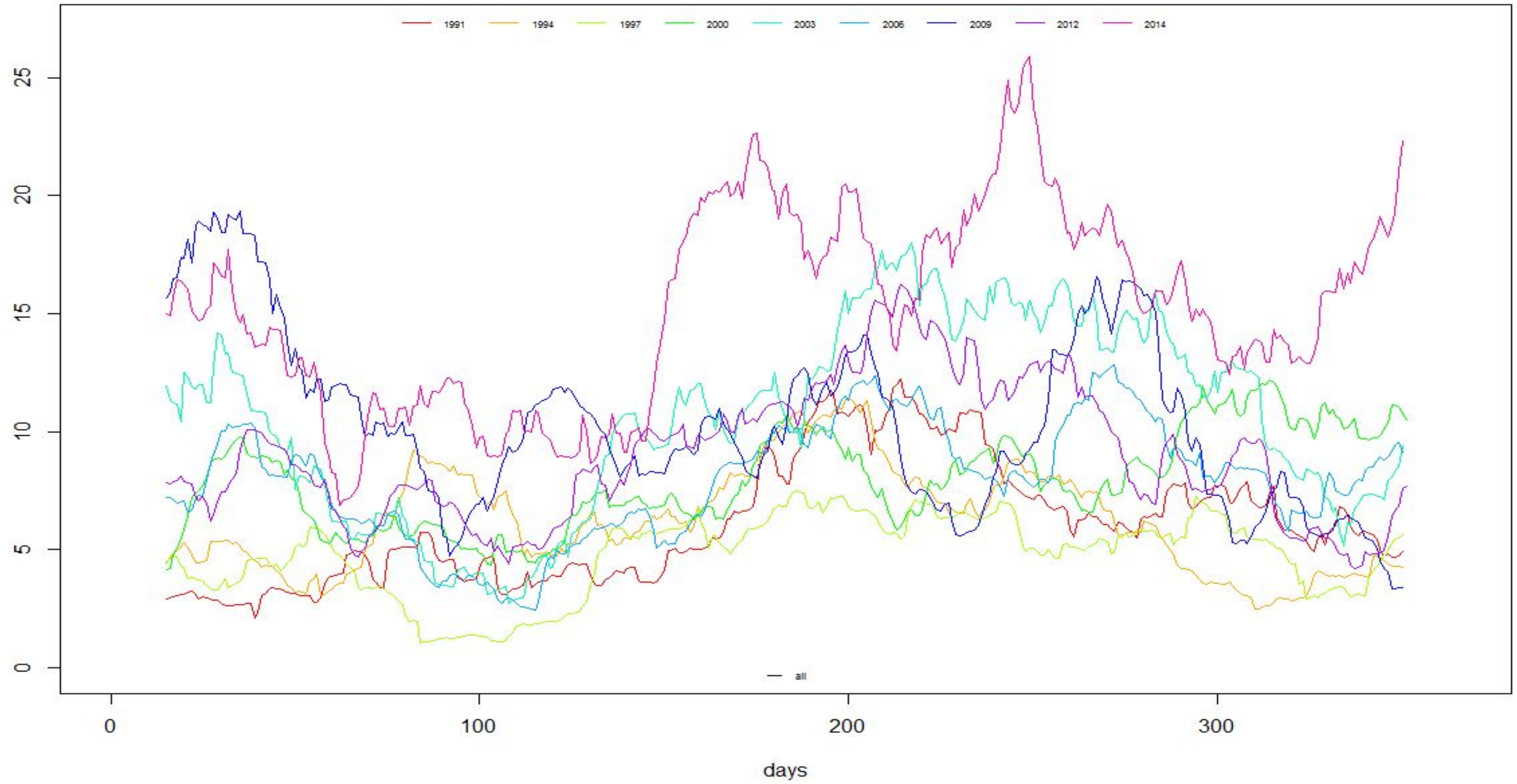
Change of temperature over the years



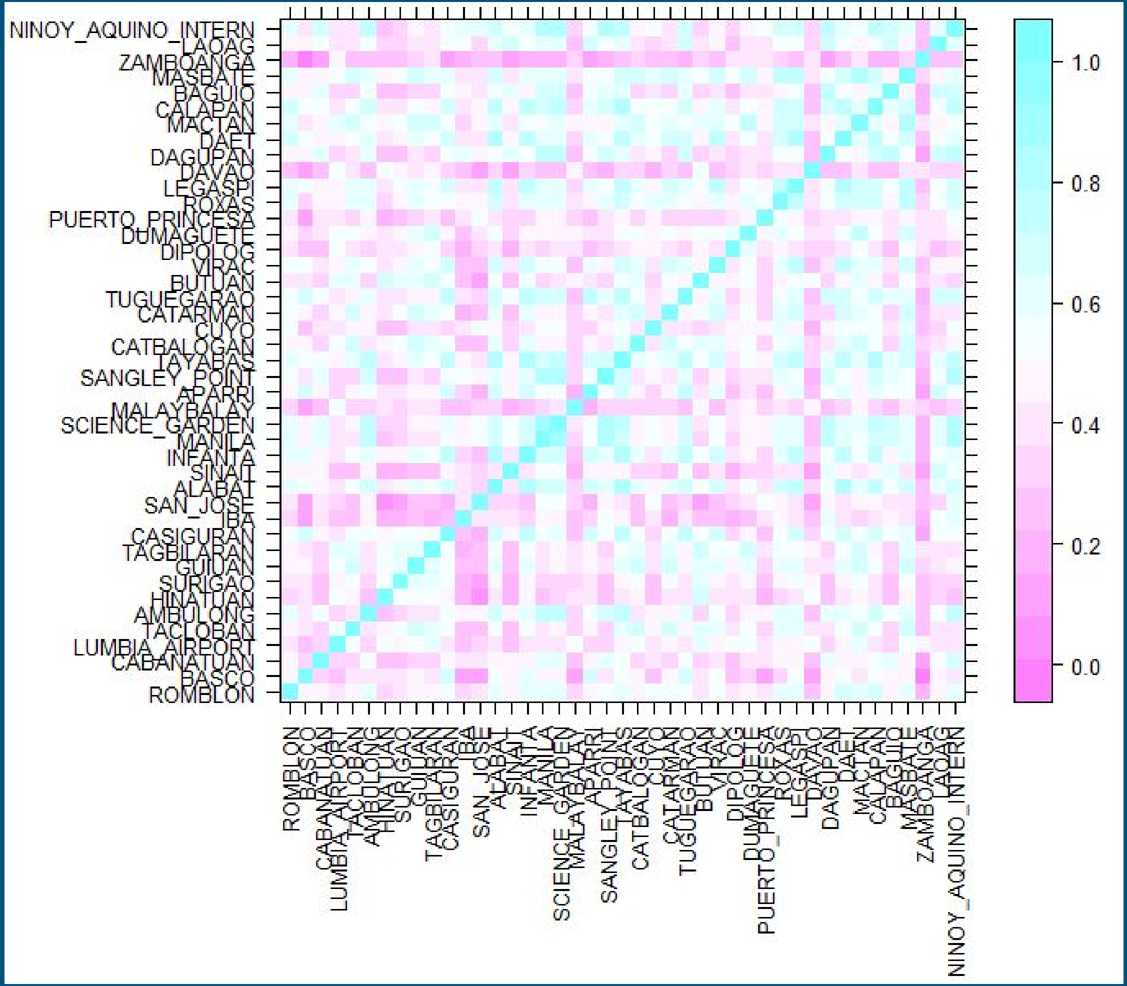
1991
1994
1997
2000
2003
2006
2009
2012
2014
all
30
29
28
27
26
25
24
0
100
200
300
days


1994
2004
2014
30
29
28
27
26
25
24
all
0
100
200
300
days

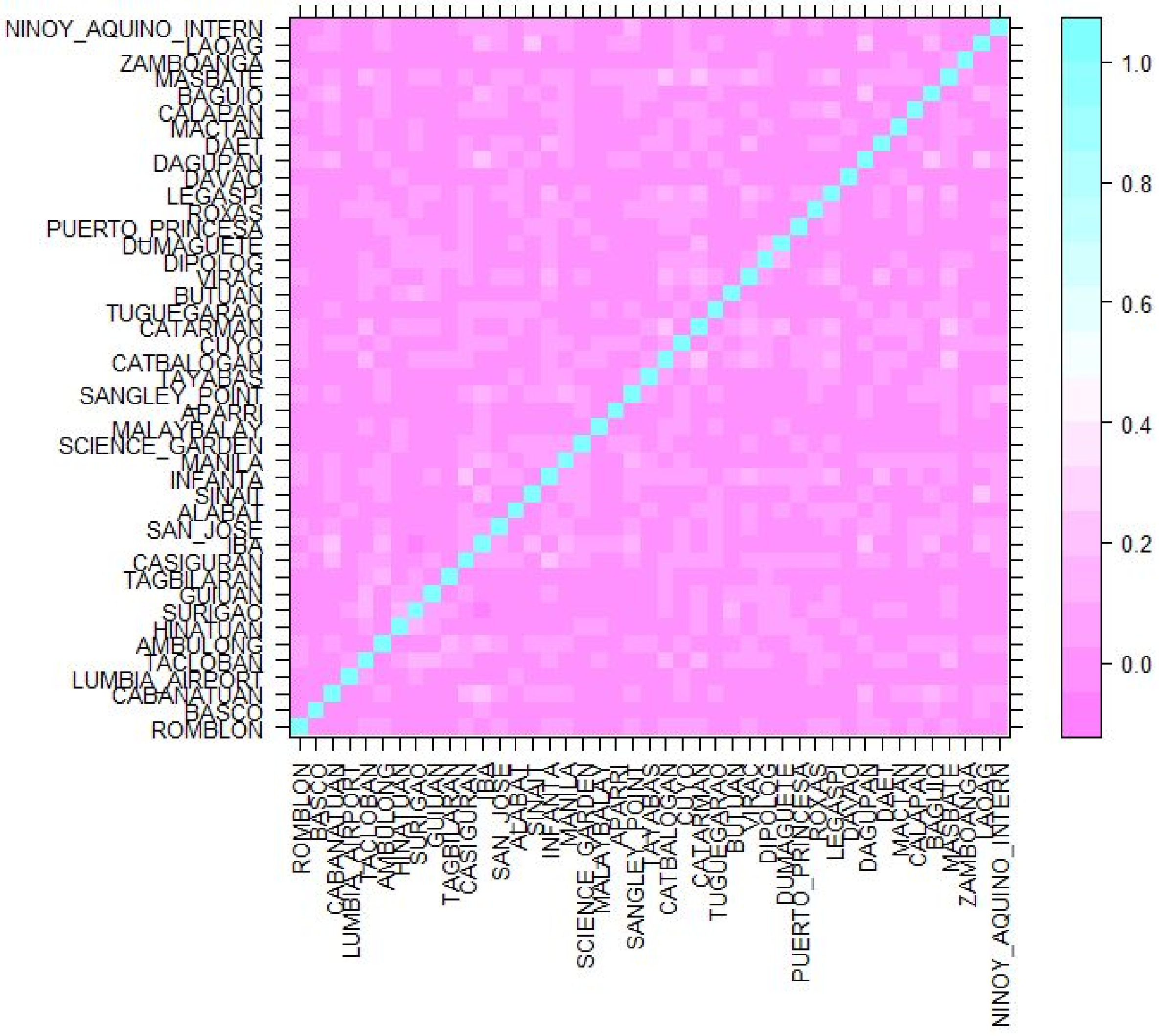
Precipitation Time Series


1991
1994
1997
2000
2003
2006
2009
2012
2014
25
20
15
10
5
0
all
0
100
200
300
days

Temperature Correlation


NINOY_AQUINO_INTERN
LAOAG
ZAMBOANGA
MASBATE
BAGUIO
CALAPAN
MACTAN
DAET
DAGUPAN
DAVAO
LEGASPI
ROXAS
PUERTO_PRINCESA
DUMAGUETE
DIPOLOG
VIRAC
BUTUAN
TUGUEGARAO
CATARMAN
CUYO
CATBALOGAN
TAYABAS
SANGLEY_POINT
APARRI
MALAYBALAY
SCIENCE_GARDEN
MANILA
INFANTA
SINAIT
ALABAT
SAN_JOSE
IBA
CASIGURAN
TAGBILARAN
GUIUAN
SURIGAO
HINATUAN
AMBULONG
TACLOBAN
LUMBIA_AIRPORT
CABANATUAN
BASCO
ROMBLON
ROMBLON
BASCO
CABANATUAN
LUMBIA_AIRPORT
TACLOBAN
AMBULONG
HINATUAN
SURIGAO
GUIUAN
TAGBILARAN
CASIGURAN
IBA
SAN_JOSE
ALABAT
SINAIT
INFANTA
MANILA
SCIENCE_GARDEN
MALAYBALAY
APARRI
SANGLEY_POINT
TAYABAS
CATBALOGAN
CUYO
CATARMAN
TUGUEGARAO
BUTUAN
VIRAC
DIPOLOG
DUMAGUETE
PUERTO_PRINCESA
ROXAS
LEGASPI
DAVAO
DAGUPAN
DAET
MACTAN
CALAPAN
BAGUIO
MASBATE
ZAMBOANGA
LAOAG
NINOY_AQUINO_INTERN
1.0
0.8
0.6
0.4
0.2
0.0

Precipitation Correlation


NINOY_AQUINO_INTERN
LAOAG
ZAMBOANGA
MASBATE
BAGUIO
CALAPAN
MACTAN
DAET
DAGUPAN
DAVAO
LEGASPI
ROXAS
PUERTO_PRINCESA
DUMAGUETE
DIPOLOG
VIRAC
BUTUAN
TUGUEGARAO
CATARMAN
CUYO
CATBALOGAN
TAYABAS
SANGLEY_POINT
APARRI
MALAYBALAY
SCIENCE_GARDEN
MANILA
INFANTA
SINAIT
ALABAT
SAN_JOSE
IBA
CASIGURAN
TAGBILARAN
GUIUAN
SURIGAO
HINATUAN
AMBULONG
TACLOBAN
LUMBIA_AIRPORT
CABANATUAN
BASCO
ROMBLON
ROMBLON
BASCO
CABANATUAN
LUMBIA_AIRPORT
TACLOBAN
AMBULONG
HINATUAN
SURIGAO
GUIUAN
TAGBILARAN
CASIGURAN
IBA
SAN_JOSE
ALABAT
SINAIT
INFANTA
MANILA
SCIENCE_GARDEN
MALAYBALAY
APARRI
SANGLEY_POINT
TAYABAS
CATBALOGAN
CUYO
CATARMAN
TUGUEGARAO
BUTUAN
VIRAC
DIPOLOG
DUMAGUETE
PUERTO_PRINCESA
ROXAS
LEGASPI
DAVAO
DAGUPAN
DAET
MACTAN
CALAPAN
BAGUIO
MASBATE
ZAMBOANGA
LAOAG
NINOY_AQUINO_INTERN
1.0
0.8
0.6
0.4
0.2
0.0

Variogram

We want to find out if there's:

1. Spatial continuity
2. Lag

Variogram

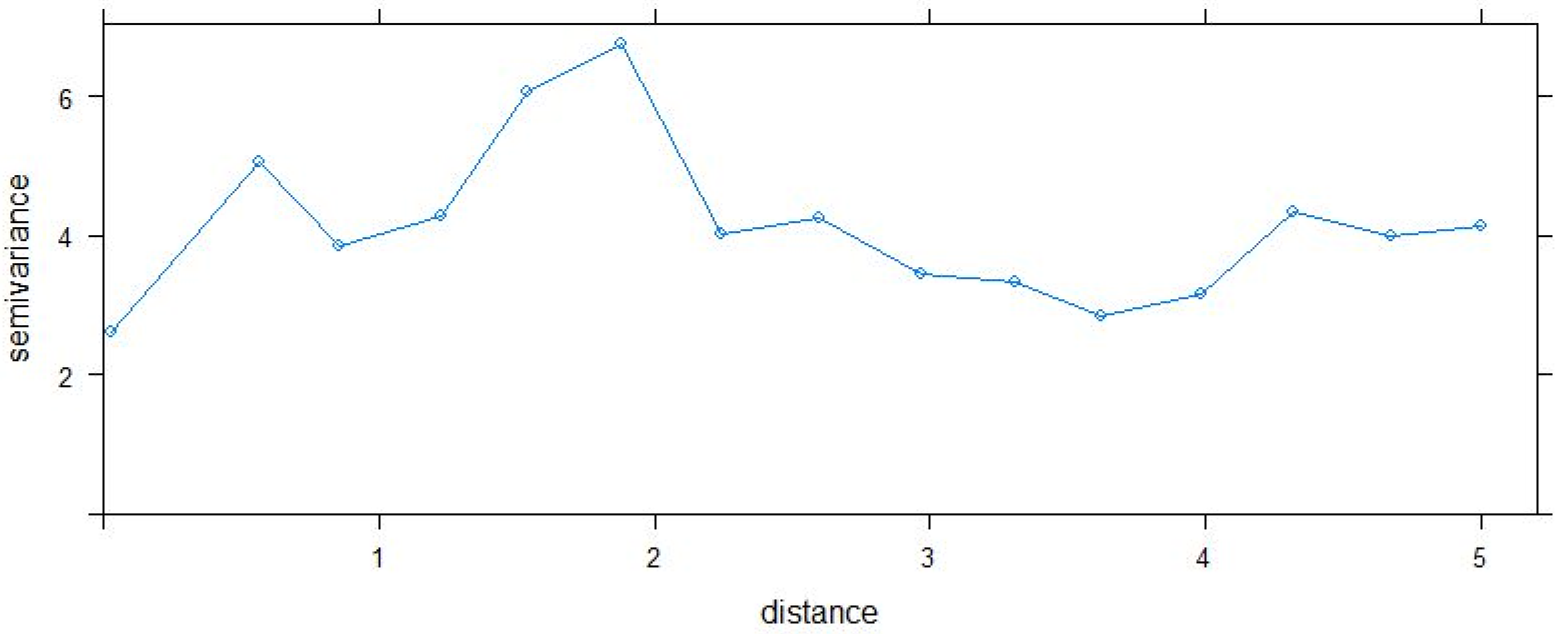
- variance vs distance
- variogram function from gstat
- 6000 samples

Variogram + fit.variogram

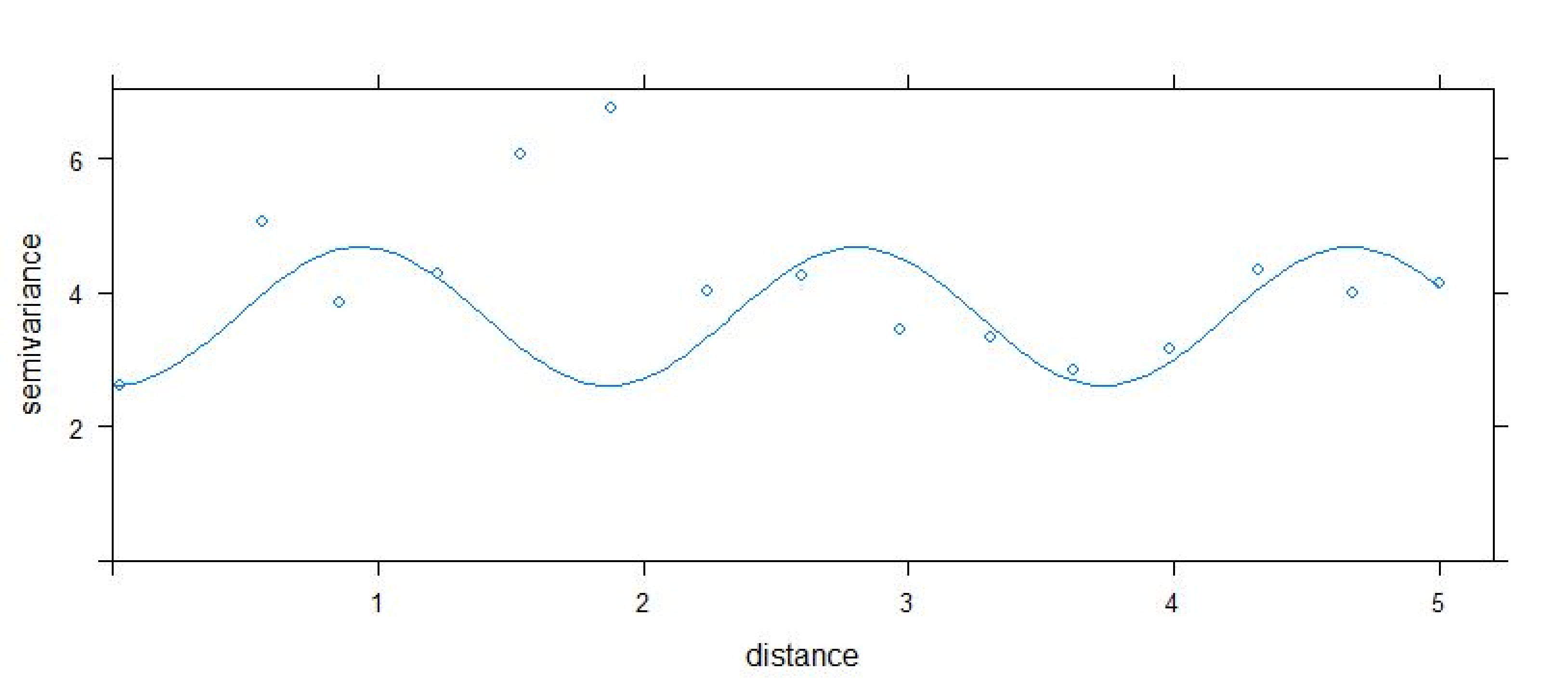
Fit.variogram

Fits the existing variogram to a model (e.g. gaussian, exponential)

Temperature (semi)variogram



Temperature (semi)variogram



Temperature (semi)variogram

Convergence error: gaussian, spherical and exponential models

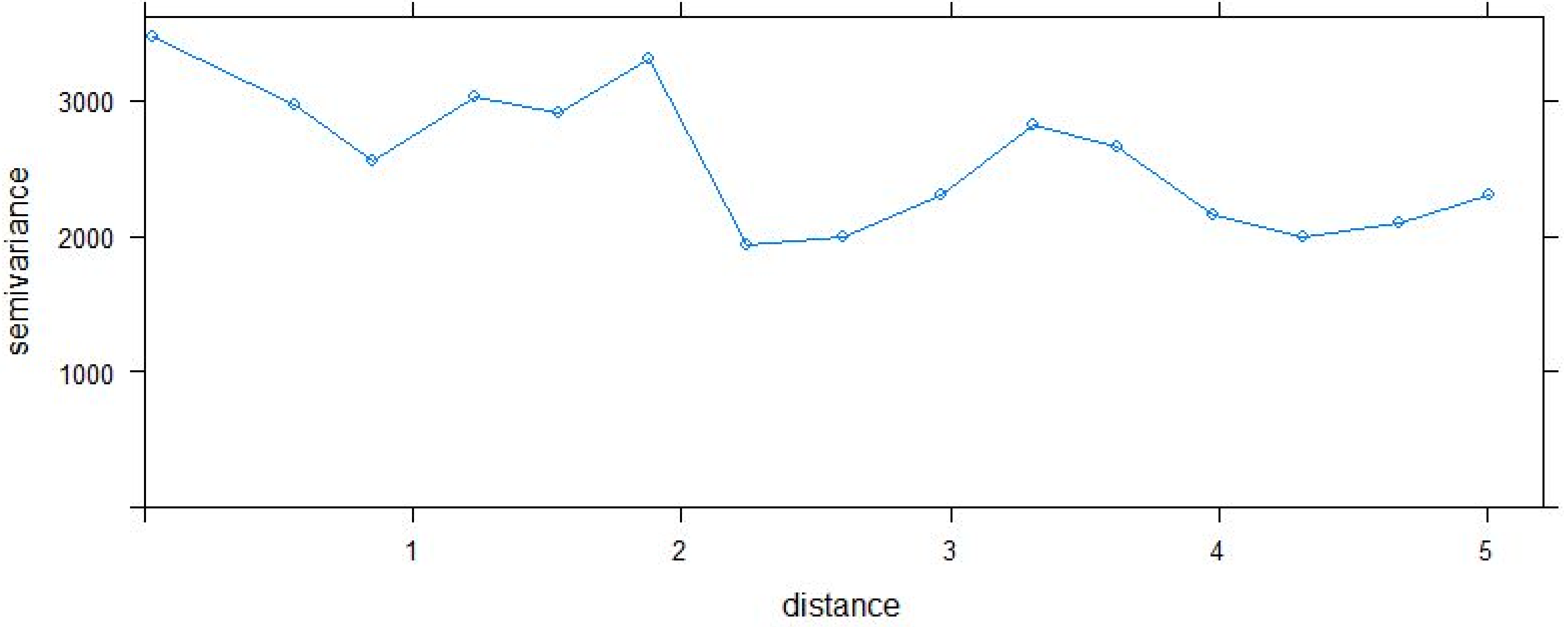
Periodic model

Initial guess: no pattern or trend

➔ Range - indicates at which distance the variogram reaches the sill value (or where it levels off)

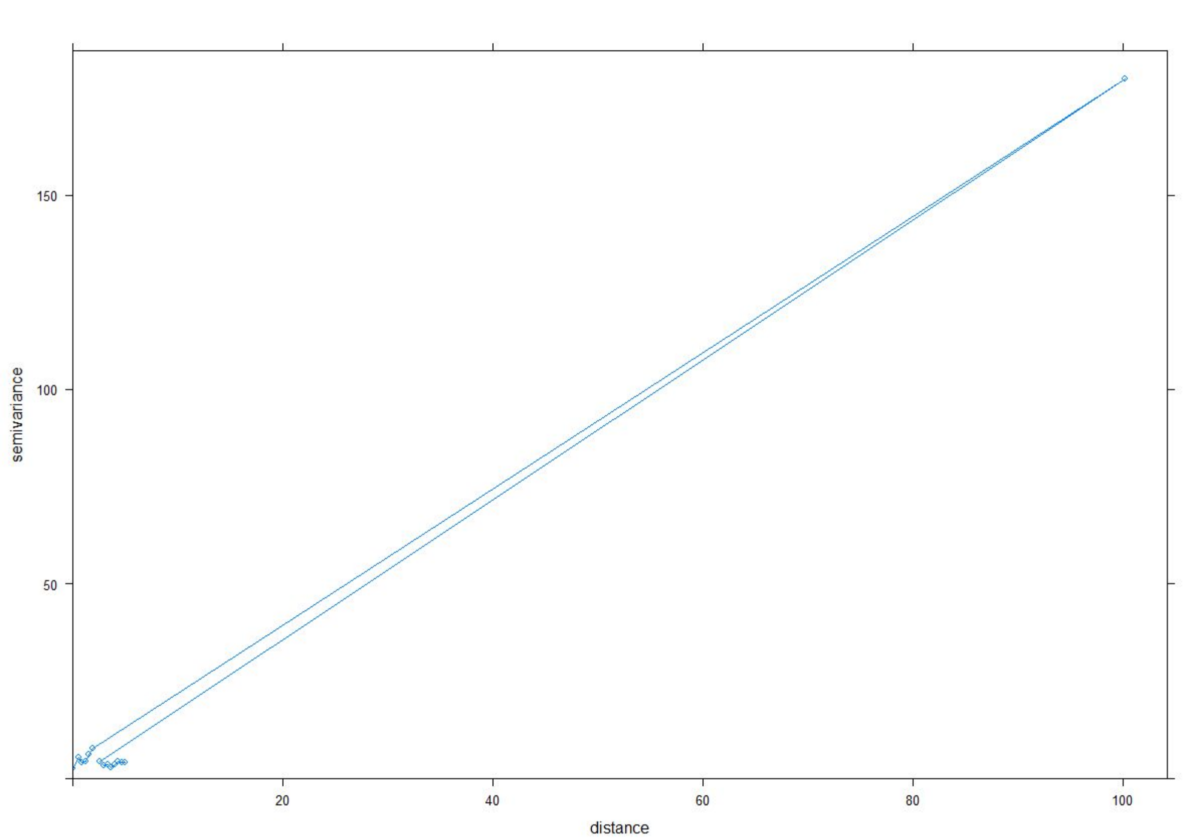
➔ Range = 0.00

Precipitation



Temperature Full Dataset

No convergence on available models



Cluster Analysis

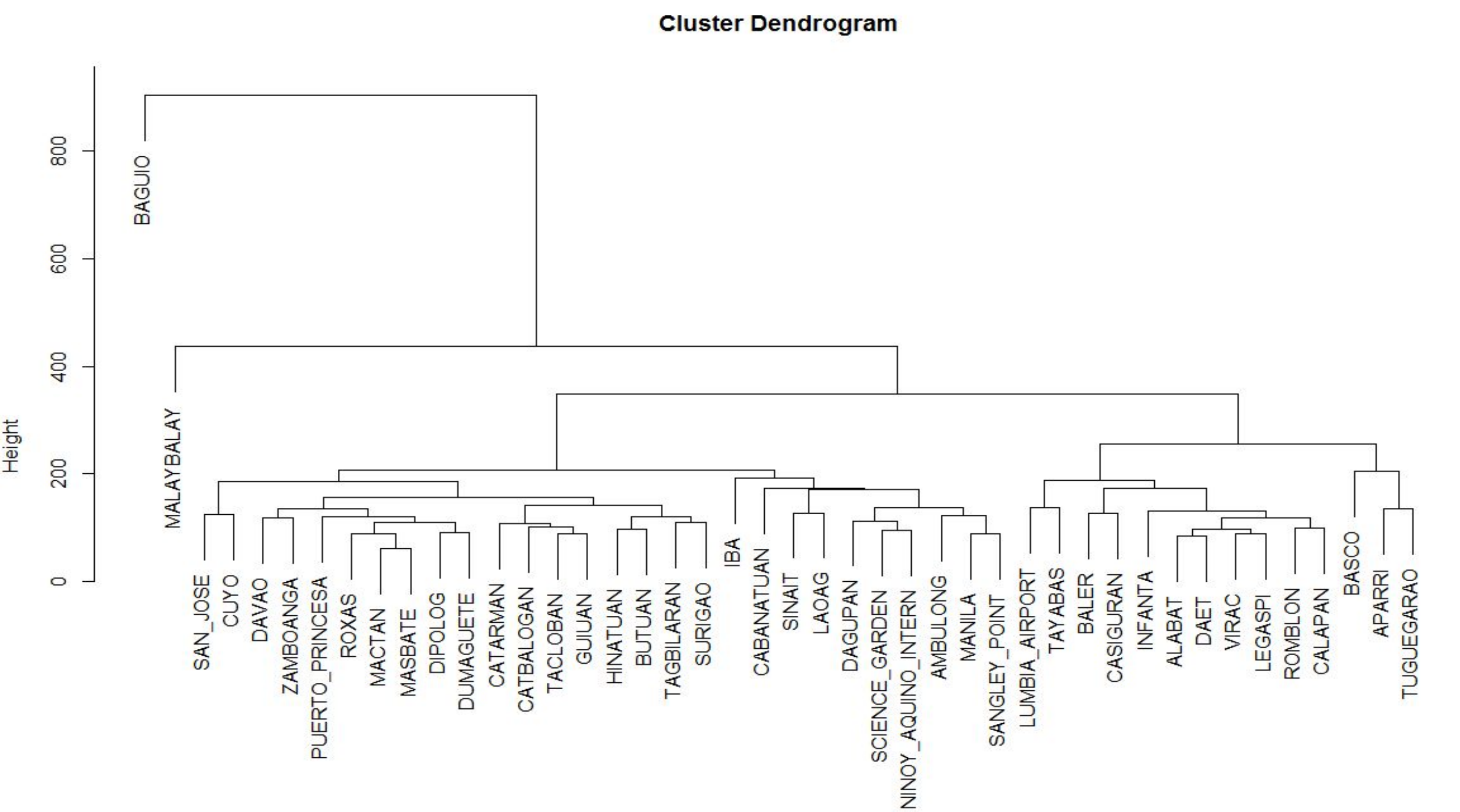
Hierarchical Clustering

“Closeness” of the stations

Distance matrix (dist) + hierarchical clustering (hclust) + split to subtrees (cutree)

Temperature, Precipitation and Temperature + Precipitation

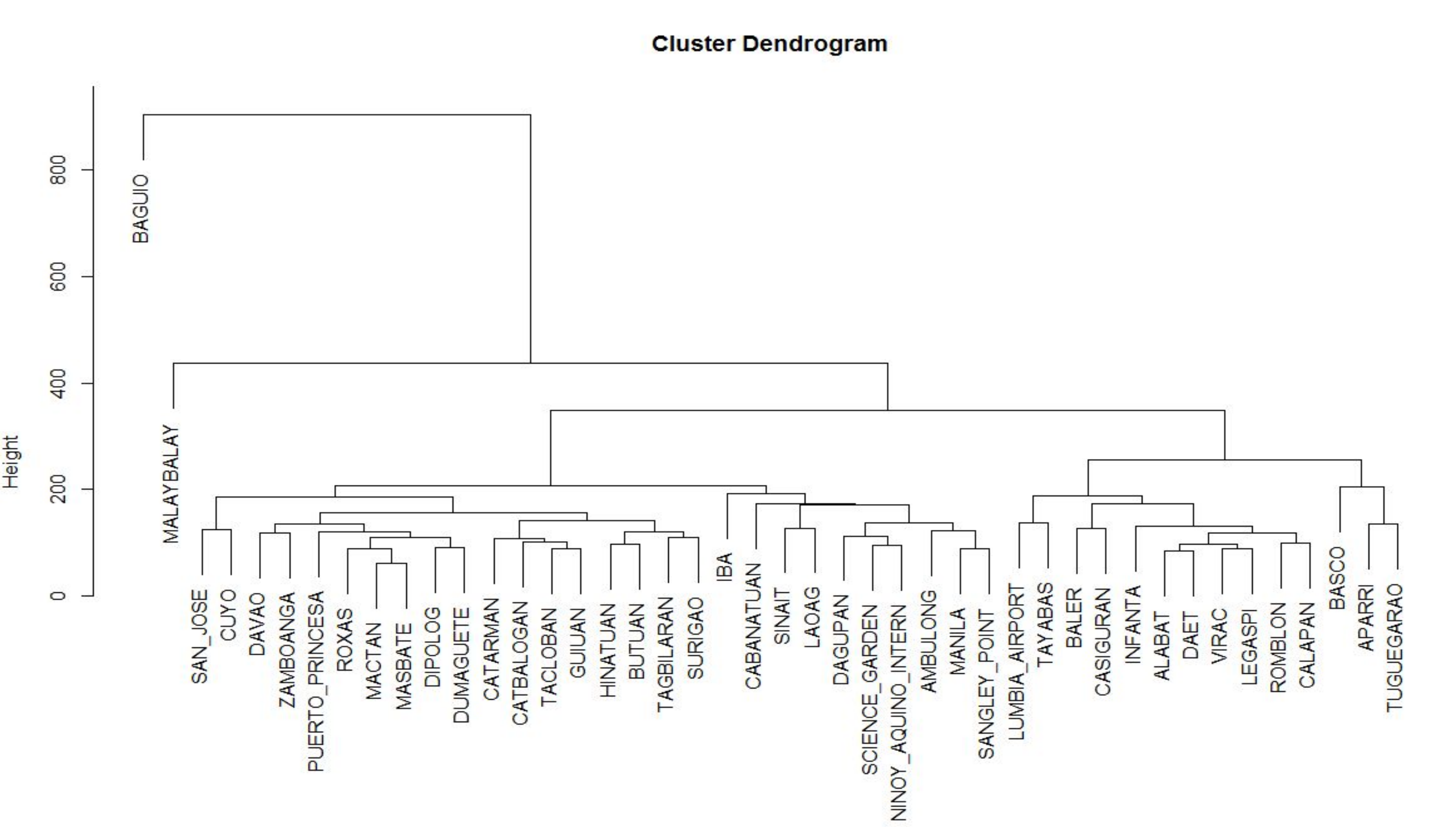
Temperature



Some observations...

- Baguio is furthest (Summer Capital)
- Cooler by 8 °C (19 °C)

Temperature



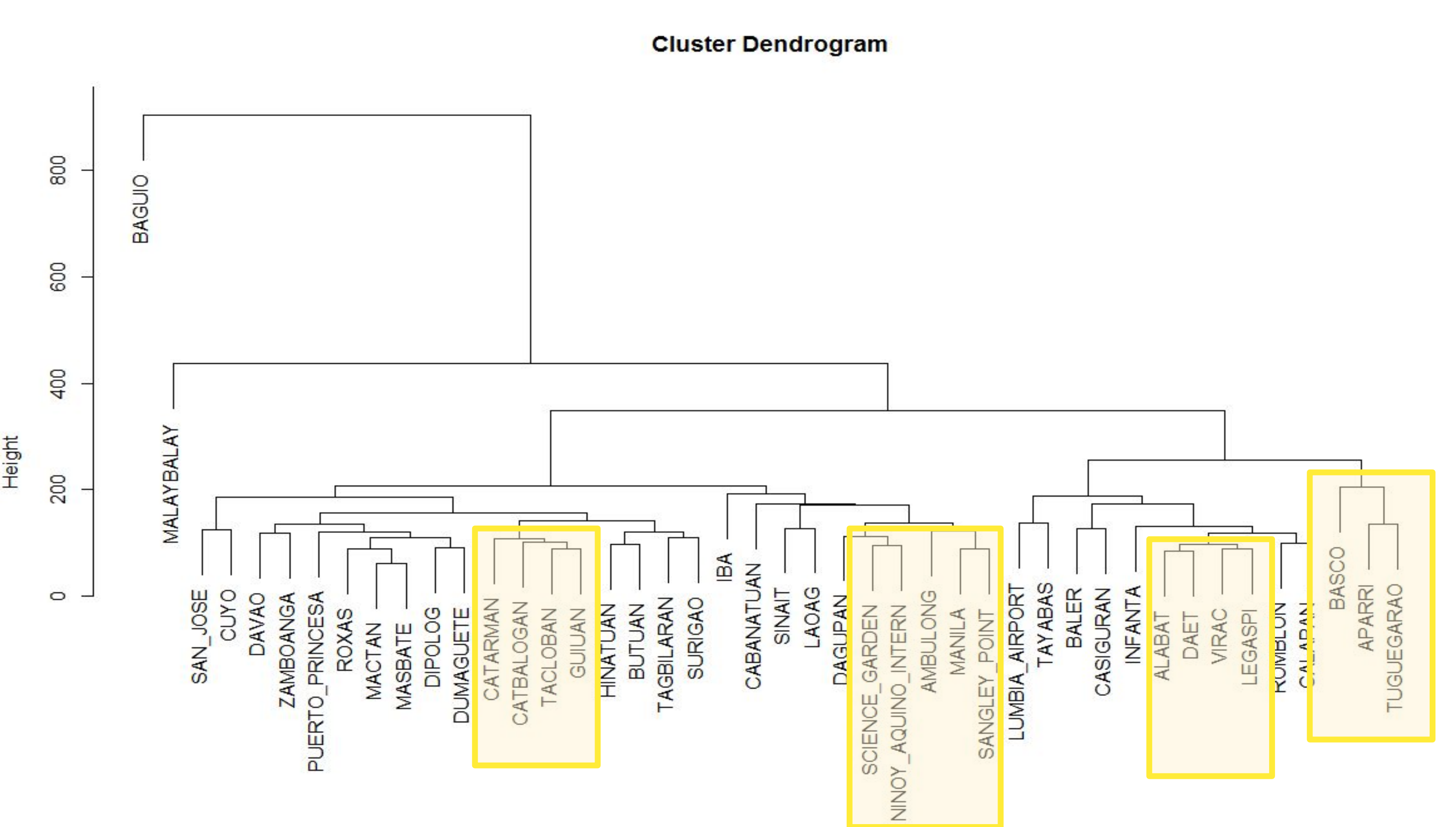
Some observations...

- **WHY, Malaybalay?? (it rhymes)**
    - 2nd to Baguio at 24 °C

“pleasant due to its altitude and the usual extreme heat of the tropical region is lacking”*

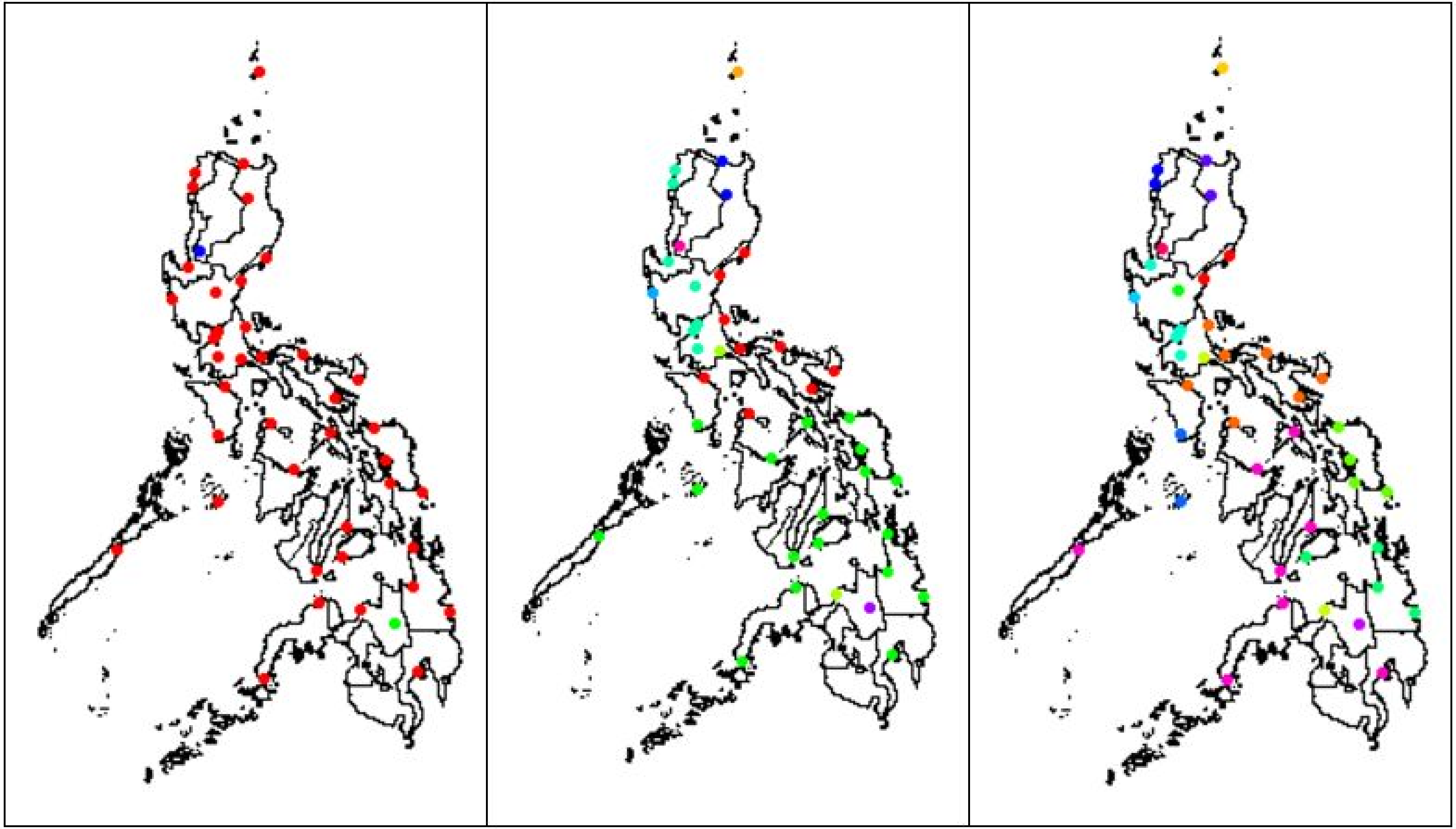
*http://www.bukidnon.gov.ph/home/index.php/about-bukidnon/general-info/climate

Temperature



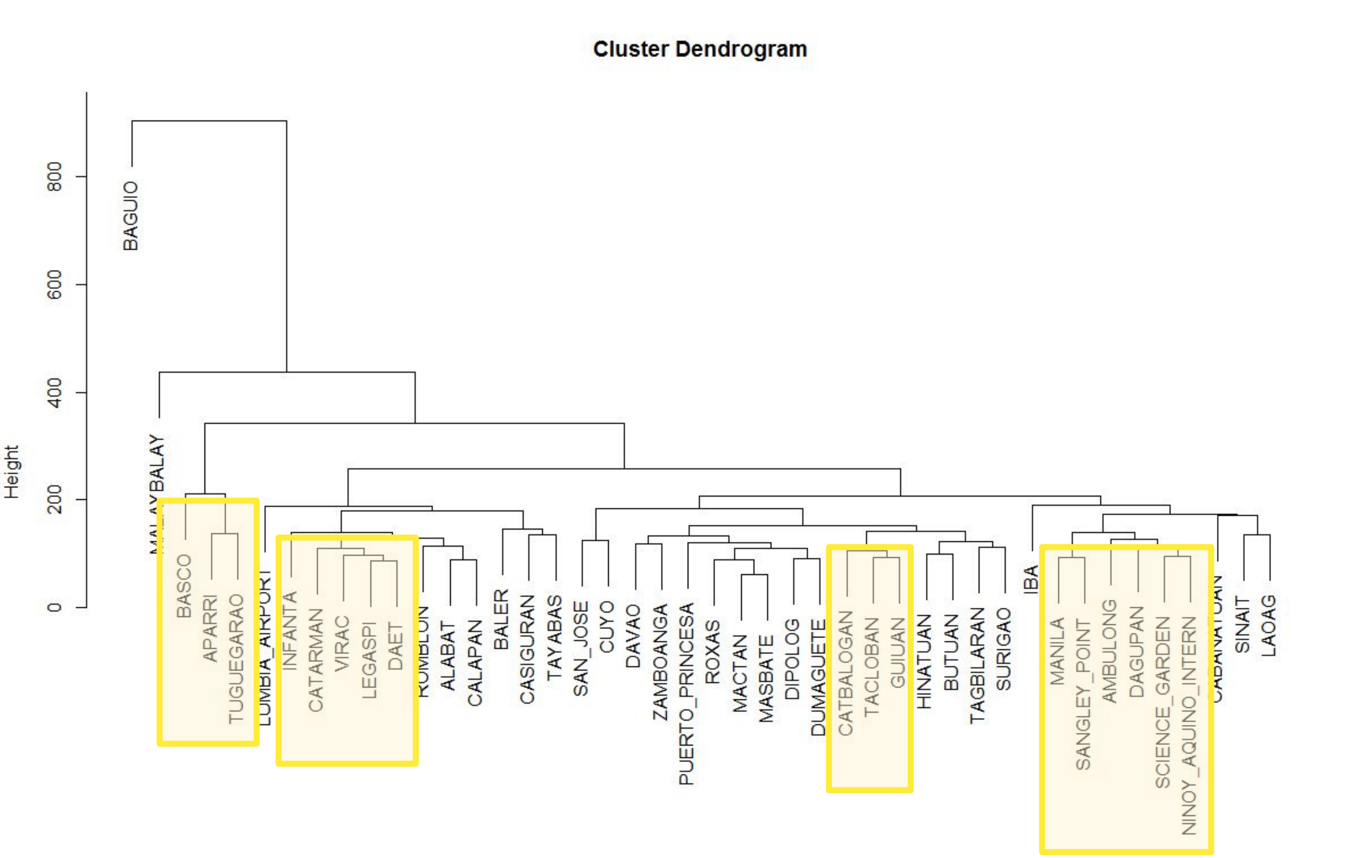
cutree

3, 9, 15



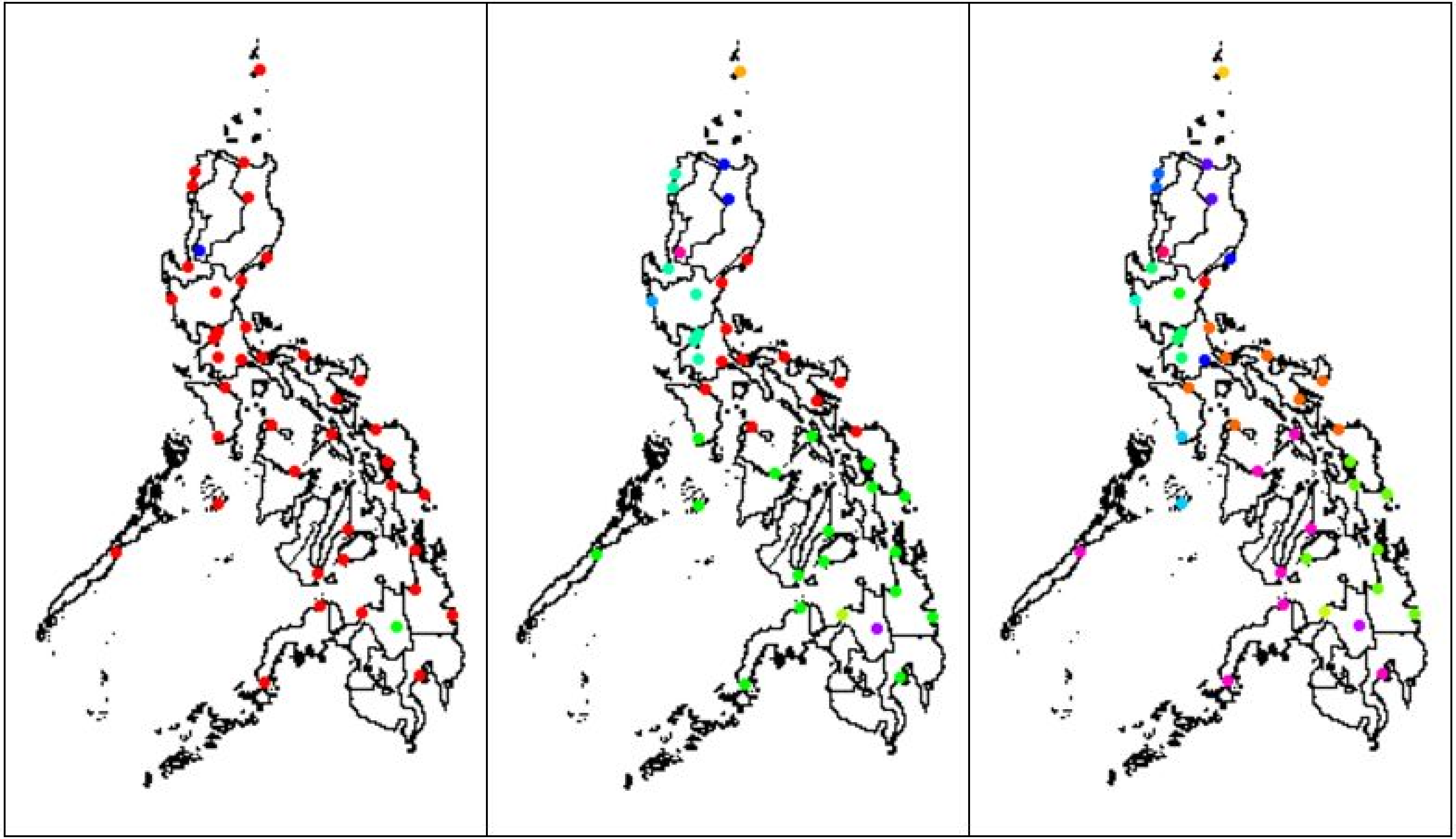
Temperature with no NAs

Replaced the NAs with the mean

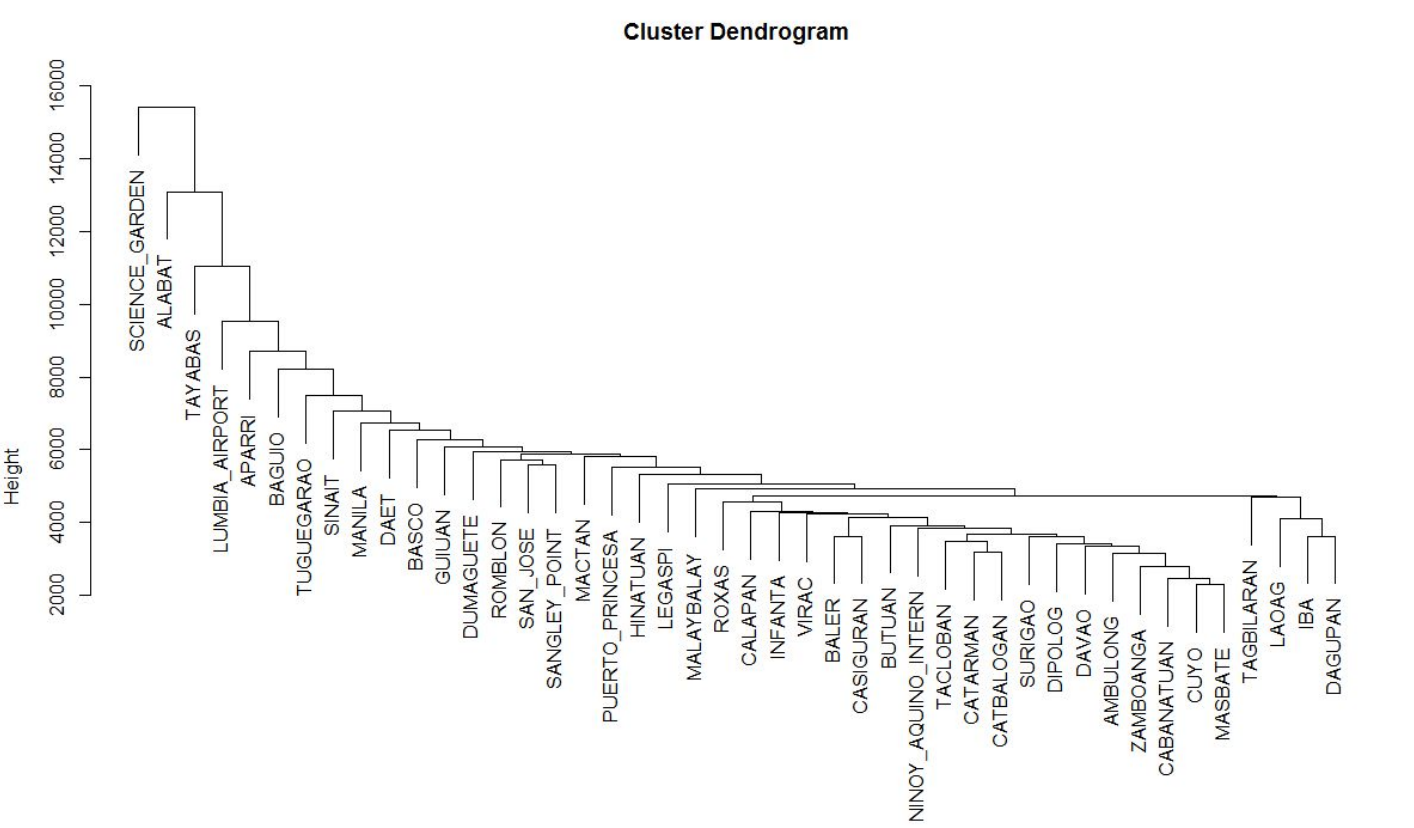


cutree

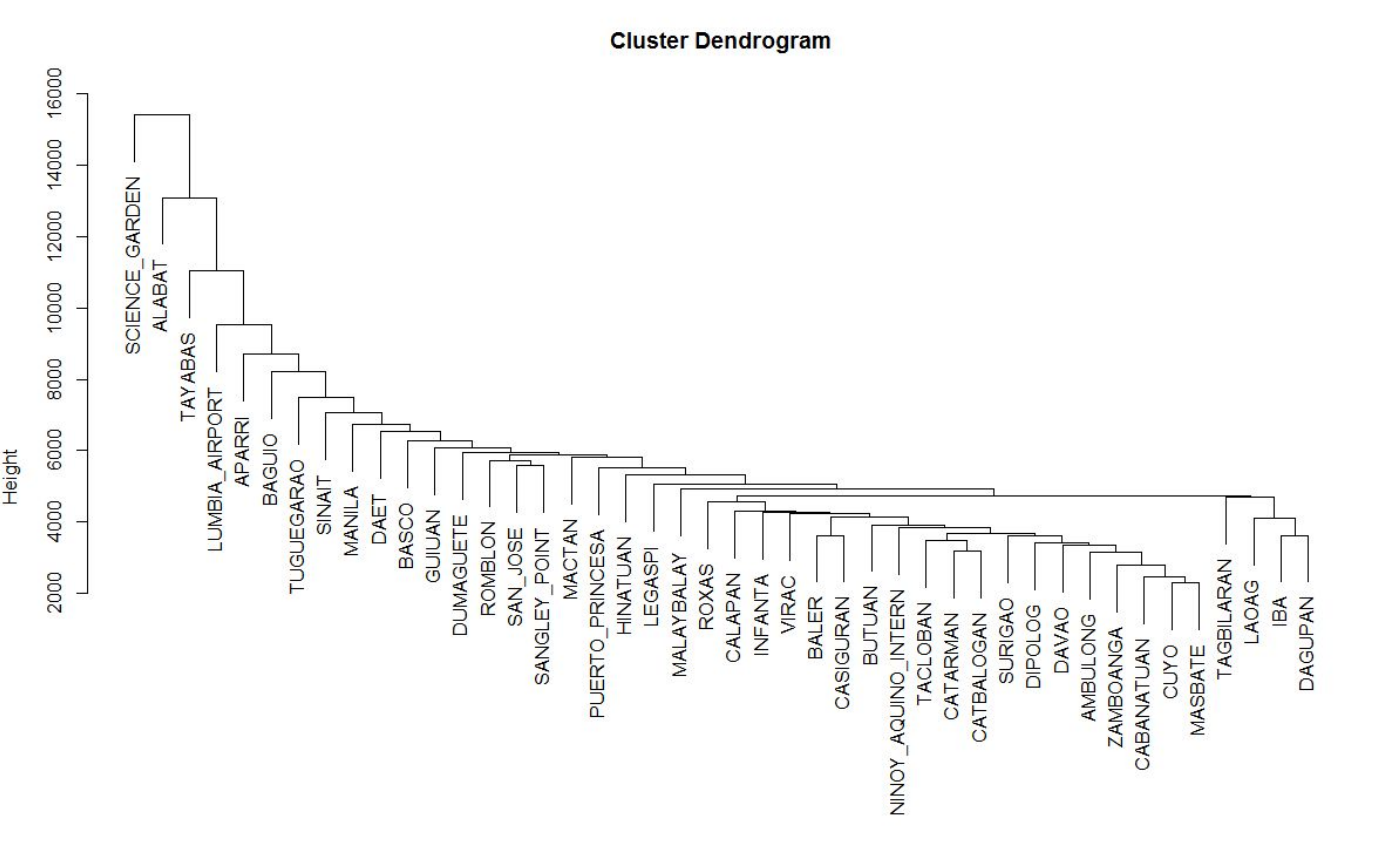
3, 9, 15



Precipitation



Precipitation with no NAs



Precipitation

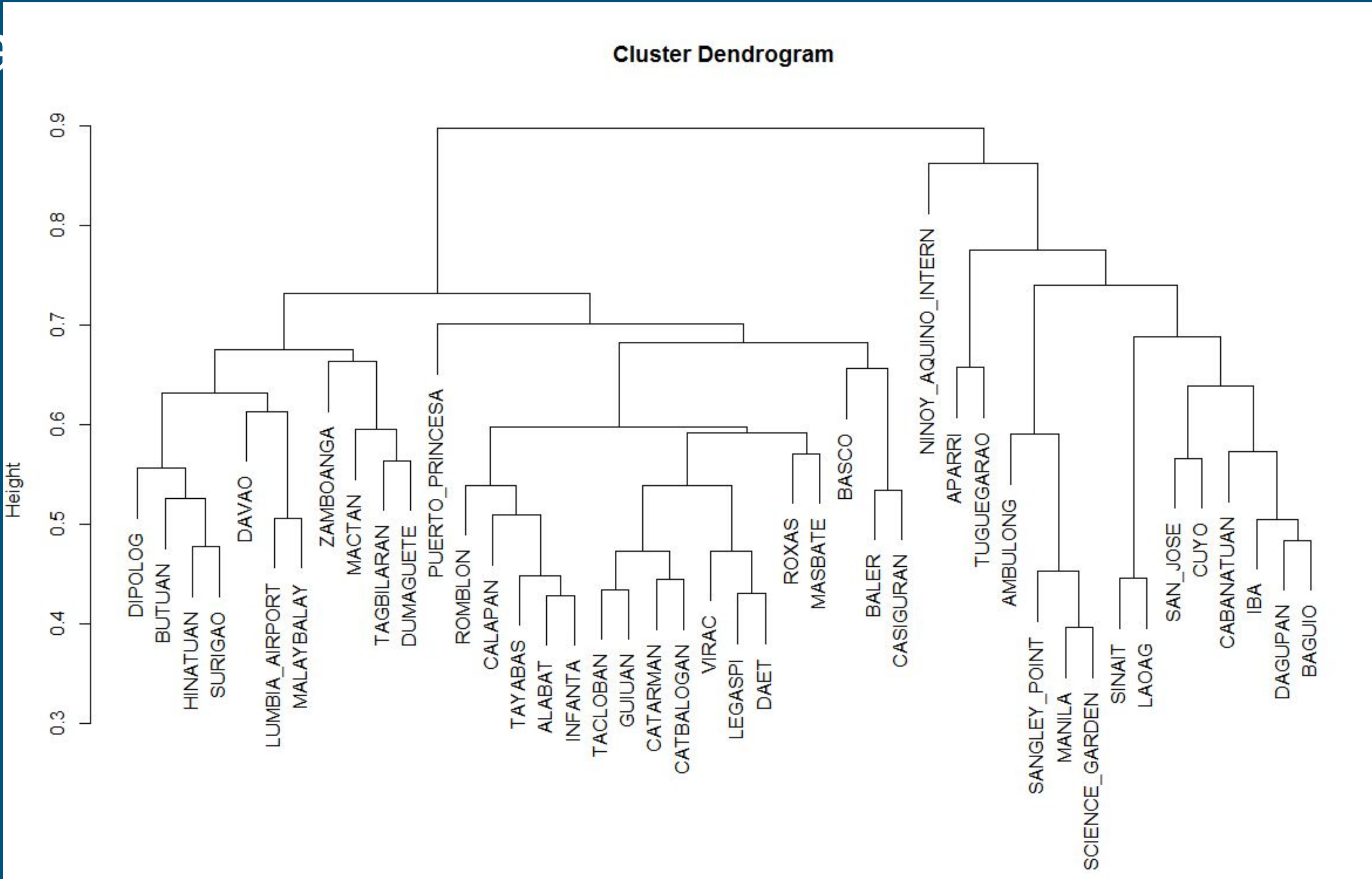
Rain or no rain

Binary distance matrix

Rain (precipitation > 0) or no rain (precipitation = 0)

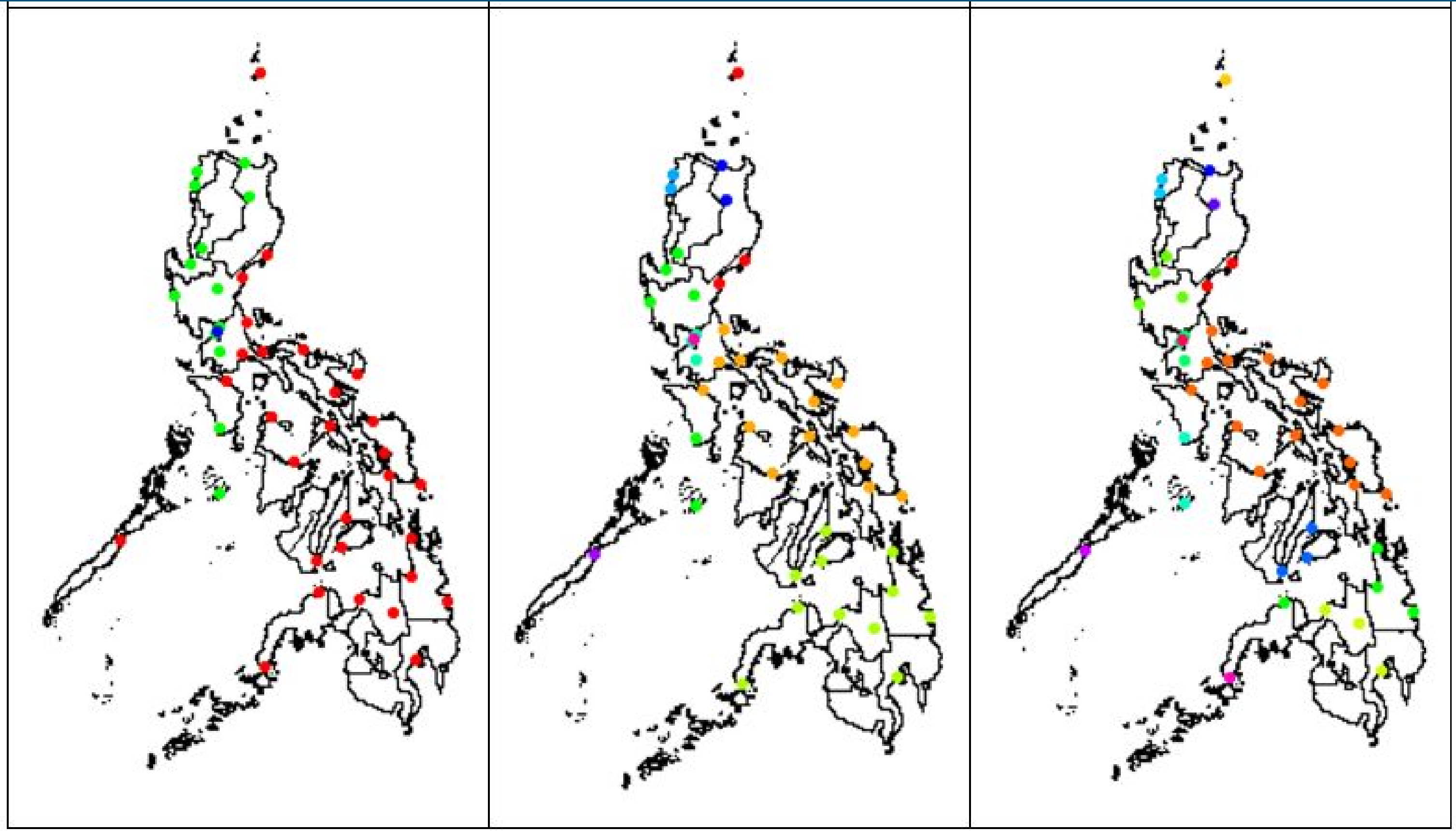
method=”binary” in the dist function

Precipita


Cluster Dendrogram
Height
0.9
0.8
0.7
0.6
0.5
0.4
0.3
DIPOLOG
BUTUAN
HINATUAN
SURIGAO
DAVAO
LUMBIA_AIRPORT
MALAYBALAY
ZAMBOANGA
MACTAN
TAGBILARAN
DUMAGUETE
PUERTO_PRINCESA
ROMBLON
CALAPAN
TAYABAS
ALABAT
INFANTA
TACLOBAN
GUIUAN
CATARMAN
CATBALOGAN
VIRAC
LEGASPI
DAET
ROXAS
MASBATE
BASCO
BALER
CASIGURAN
NINOY_AQUINO_INTERN
APARRI
TUGUEGARAO
AMBULONG
SANGLEY_POINT
MANILA
SCIENCE_GARDEN
SINAIT
LAOAG
SAN_JOSE
CUYO
CABANATUAN
IBA
DAGUPAN
BAGUIO

Precipitation

3, 9, 15

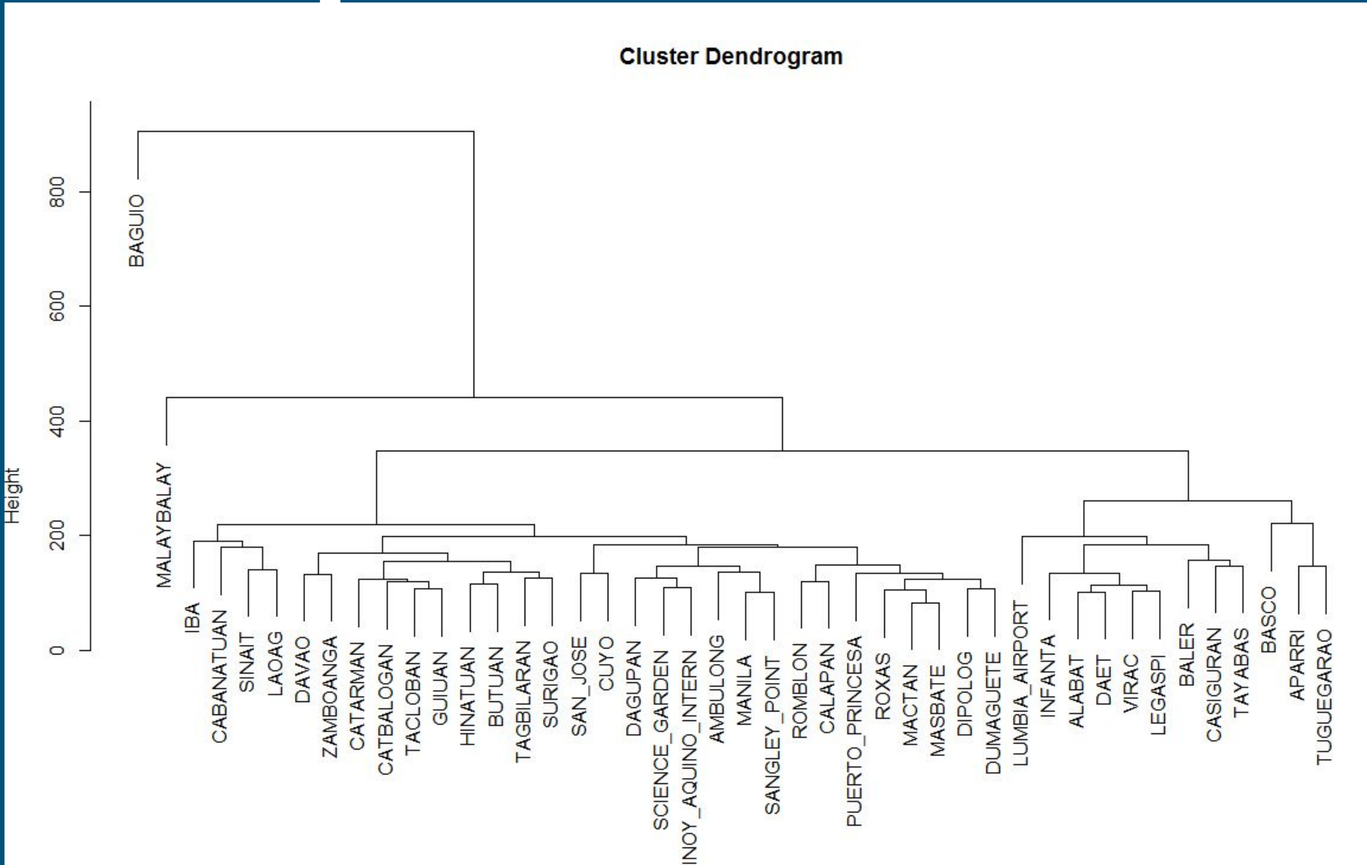


Temperature + Precipitation

Converted the precipitation to binary

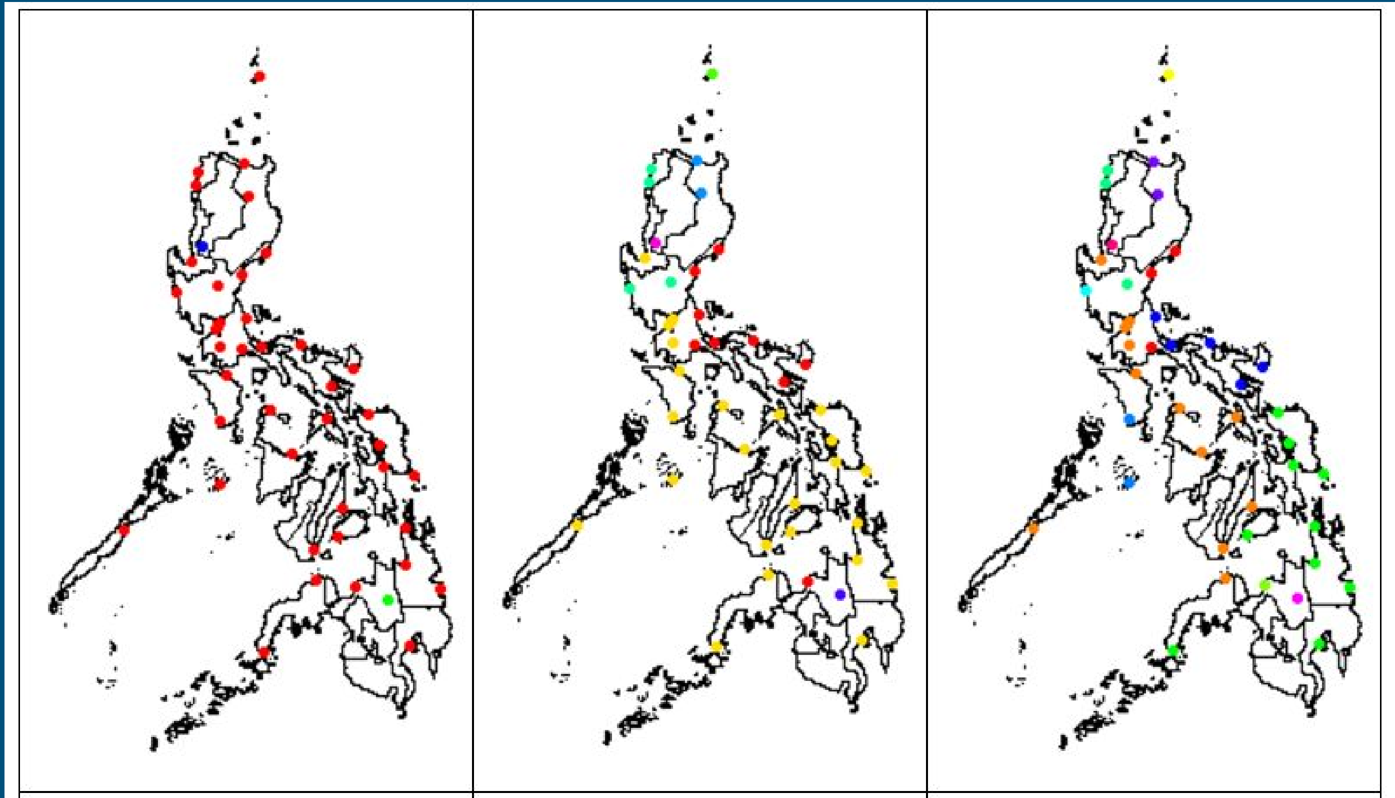
Hclust + dist (complete distance measure) + cutree

Temperature + Precipitation



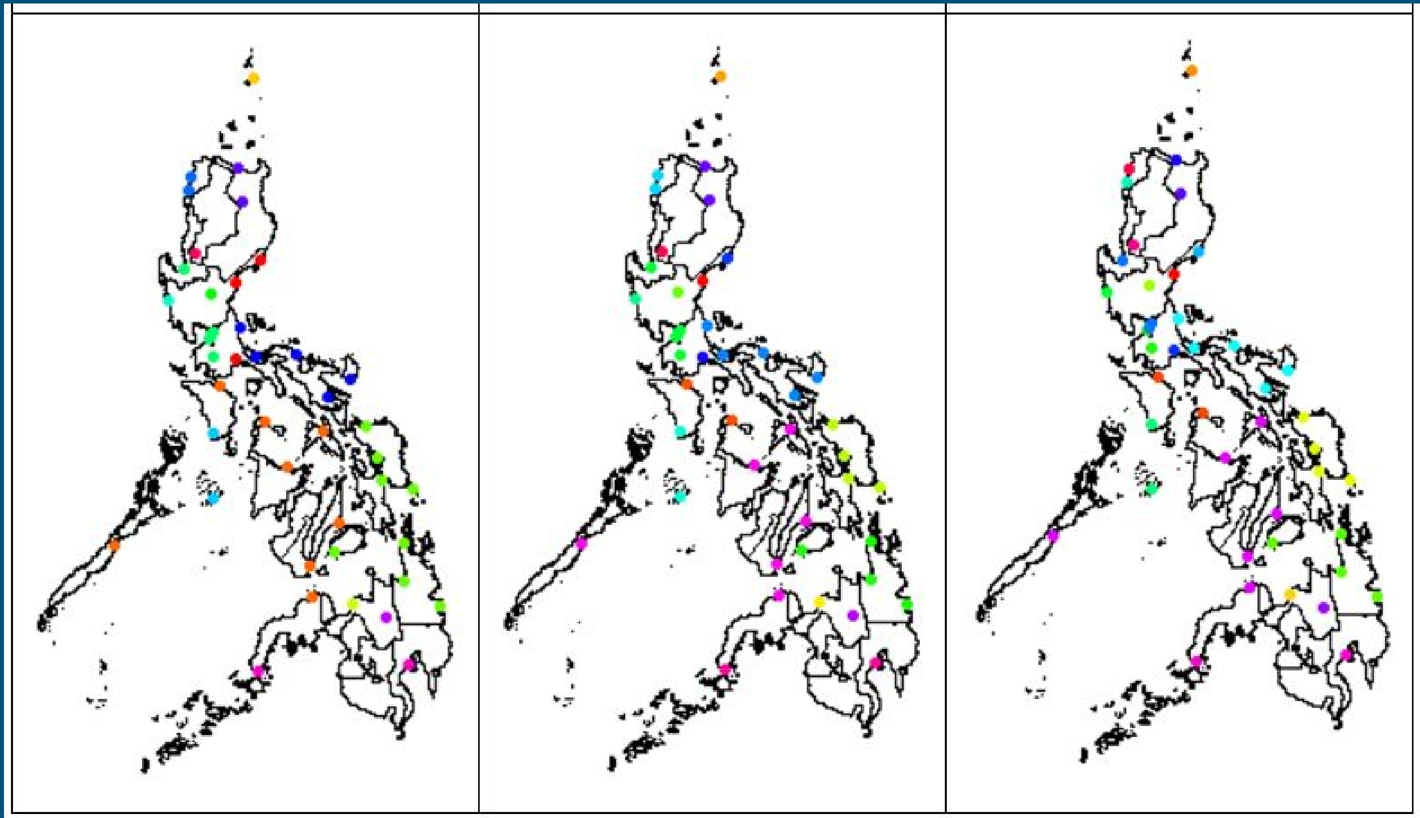
Temperature + Precipitation

3, 7, 12



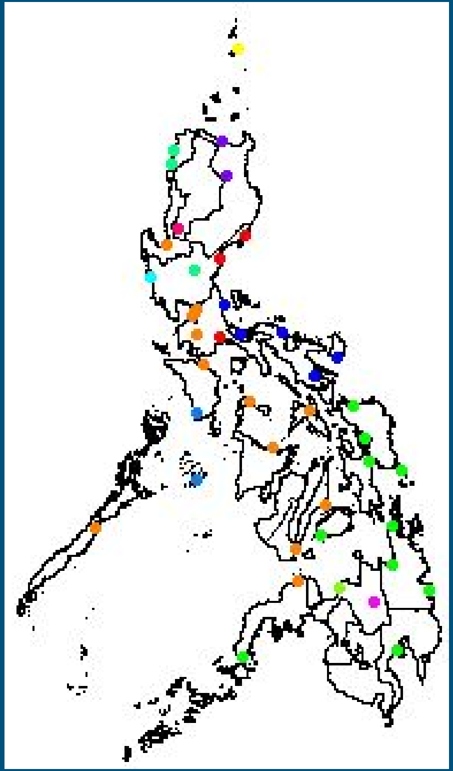
Temperature + Precipitation

15, 19, 22



Some observations...

- Eastern coast is “further” from the western coast
    - Bicol region and Quezon province
    - Samar and Leyte



Thank you