필요 소프트웨어 설치 및 설정

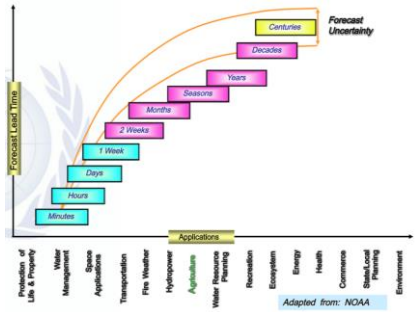
- R: https://cran.r-project.org/bin/windows/base/
- Rtools: https://cran.r-project.org/bin/windows/Rtools/
- R-Studio: https://www.rstudio.com/products/RStudio/
- QGIS: https://qgis.org/ko/site/
- Panoply: https://www.giss.nasa.gov/tools/panoply/download/

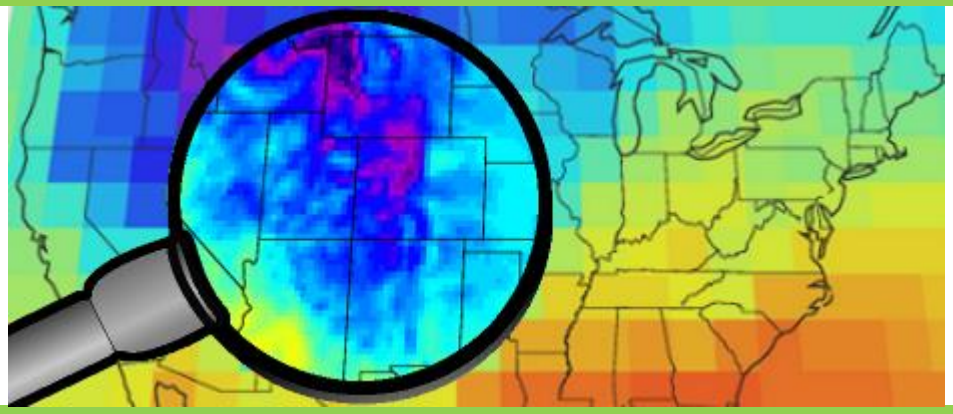
You've already installed all required software from the previous session for

1. Gridded proxy data extraction

APCC
APEC CLIMATE CENTER

Climate data QC and Gap Filling



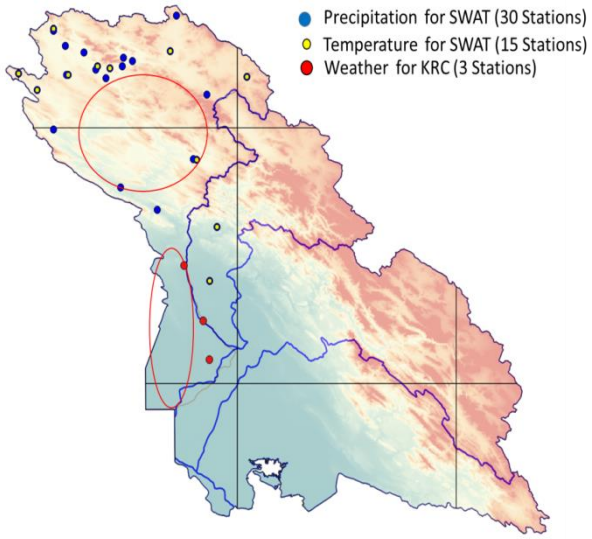


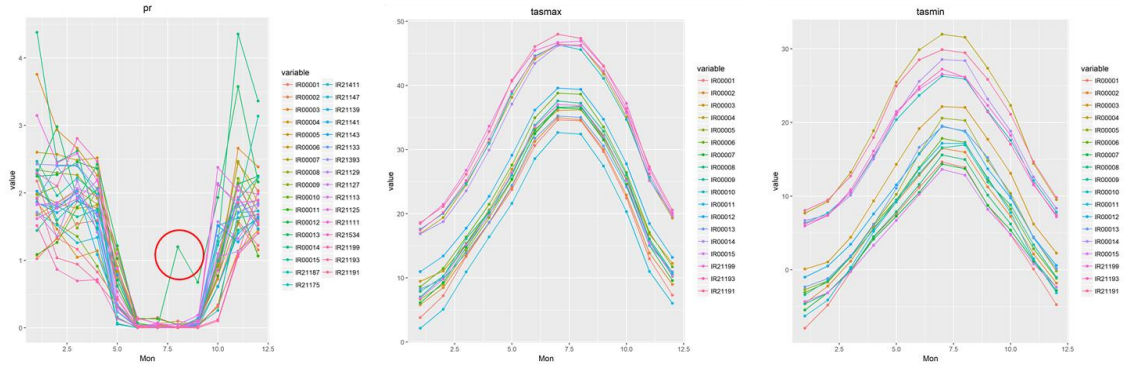
Jaepil Cho & Kwang-Hyung Kim

2018/10/16

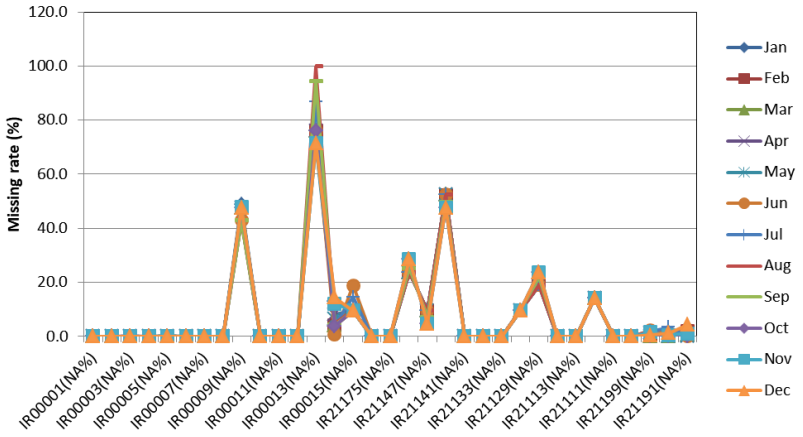
Table of contents

Example of weather stations with missing data





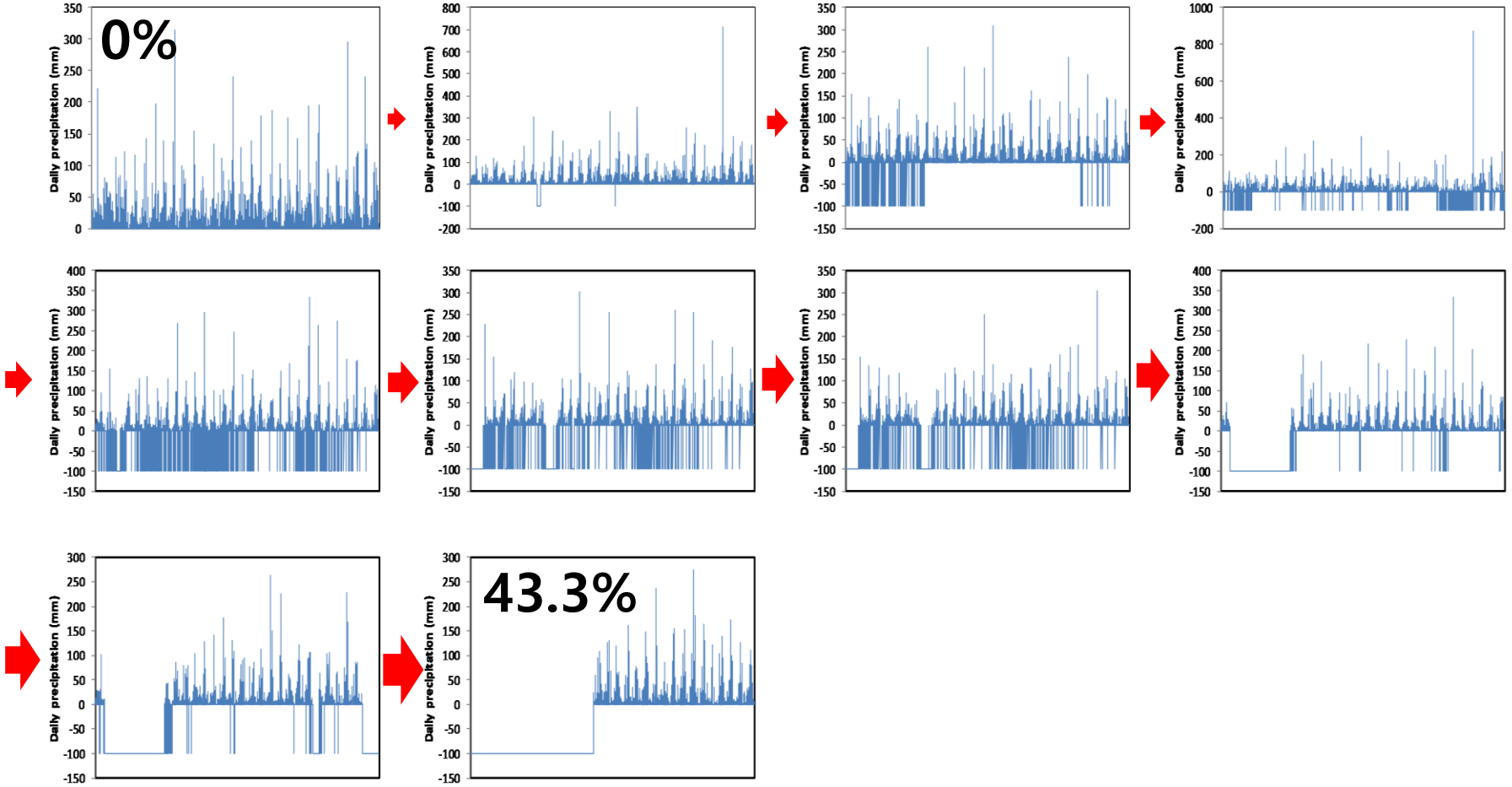
(a)



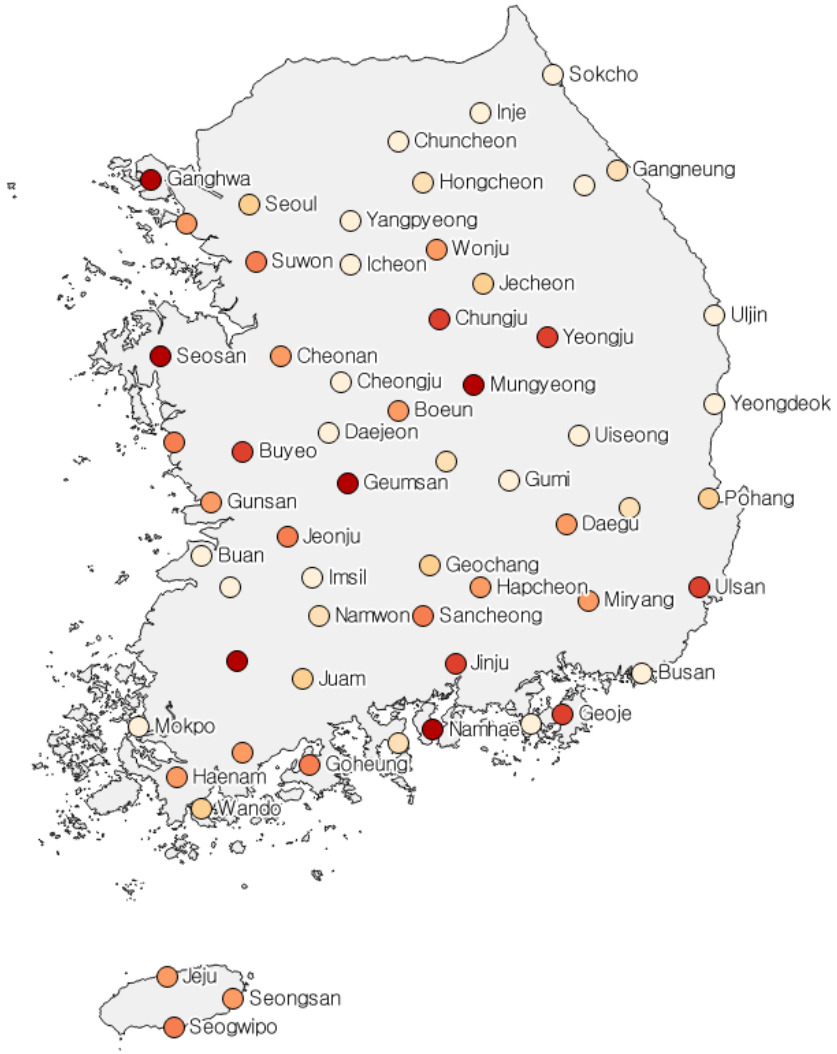
(b)

Fig. 3. Comparison of (a) monthly average of precipitation, maximum temperature, and minimum temperature and (b) missing rate of precipitation stations.

Perfect observed data with various ratio of artificial missing

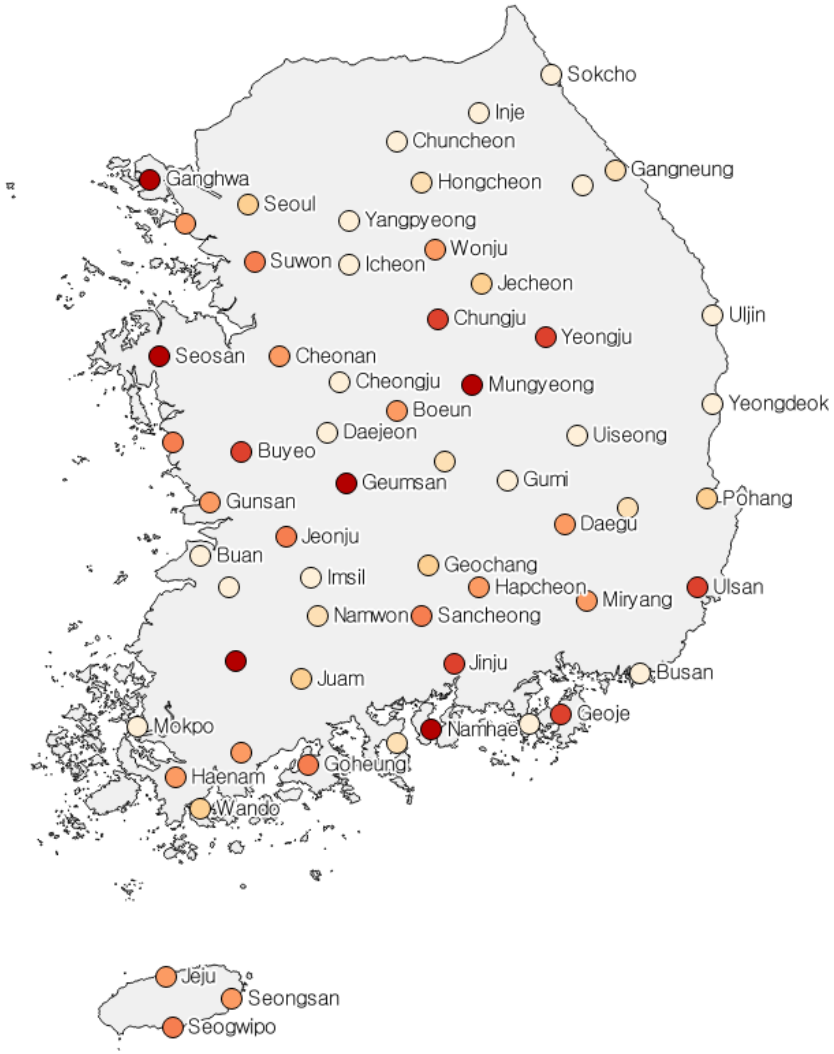


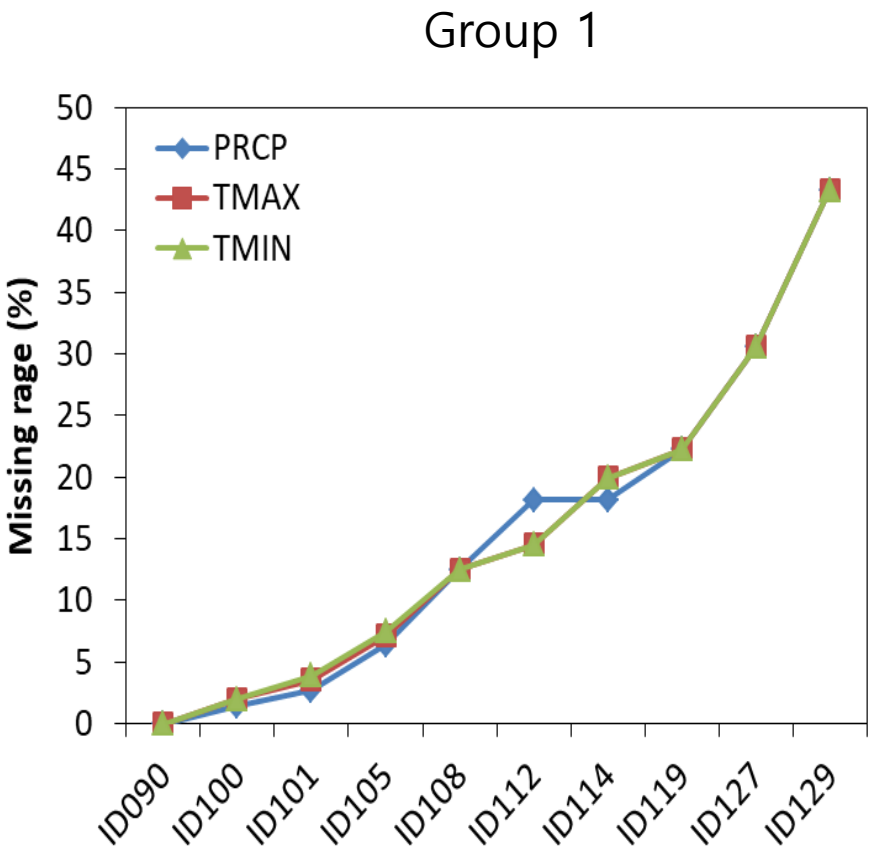
Example of KMA weather station data



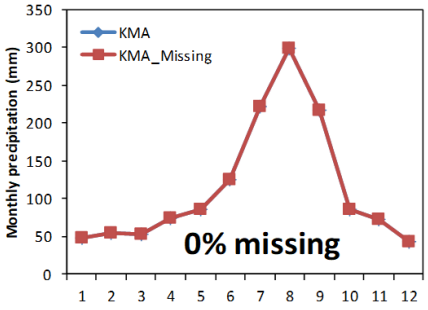
Group	Weather stations with increasing ratio of missing
1	Sokcho→Daekwanryeong→Chuncheon→Gangneung→Seoul→Incheon→Wonju→Suwon→Chungju→Seosan
2	울진→청주→대전→추풍령→포항→군산→대구→전주→울산→광주
3	부산→통영→목포→여수→완도→제주→성산→서귀포→진주→강화
4	양평→인천→인제→홍천→제천→보은→천안→보령→부여→금산
5	부안→임실→정읍→남원→주암→장흥→해남→고흥→영주→문경
6	영덕→의성→구미→영천→고창→합천→밀양→산청→ 거제→ 남해

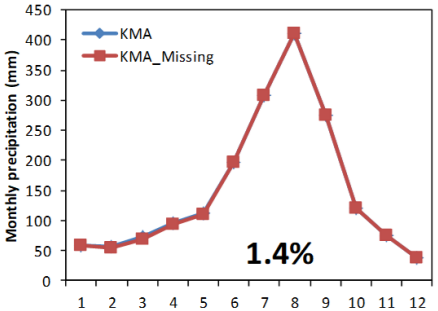
Group 1 observed data with various ratio of artificial missing

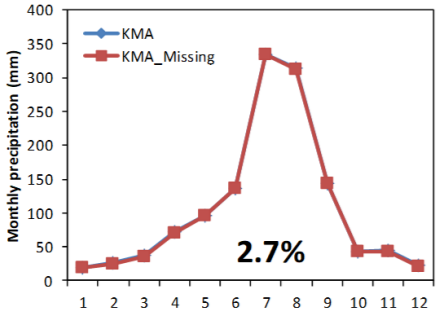


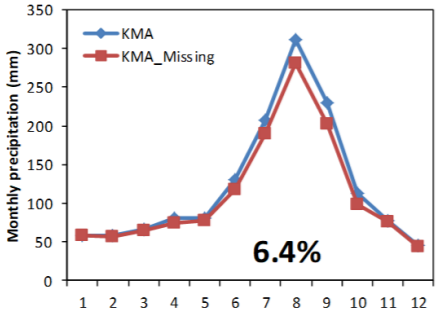


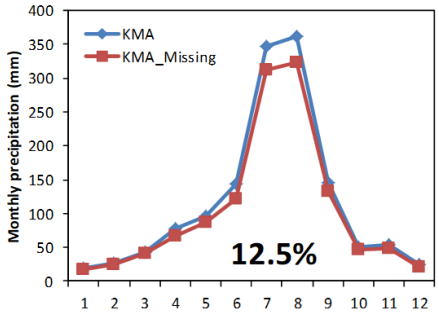
Monthly precipitation (30-yrs) climatology with increased missing

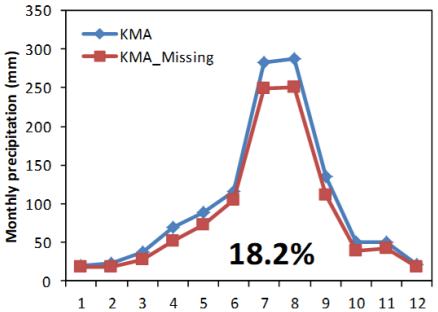


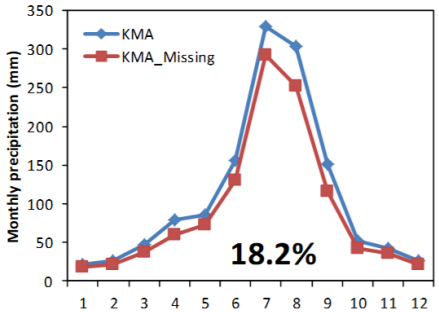


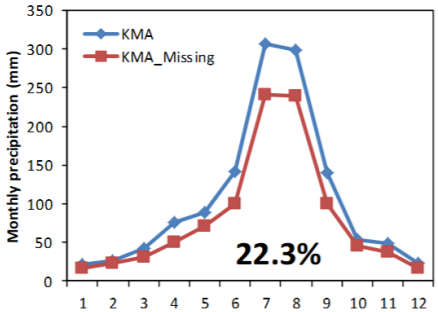


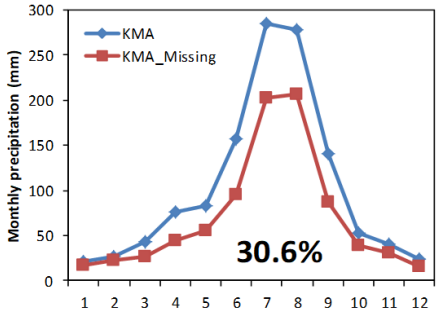


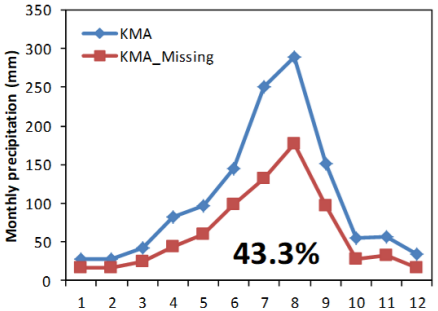




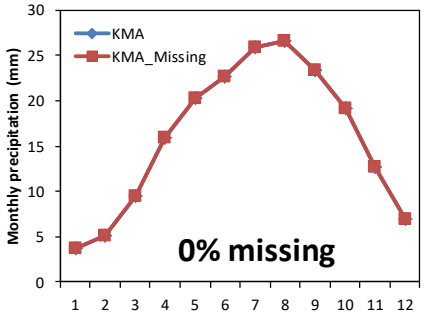


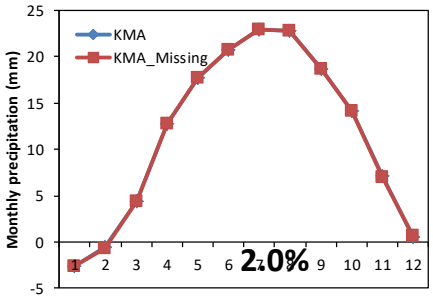


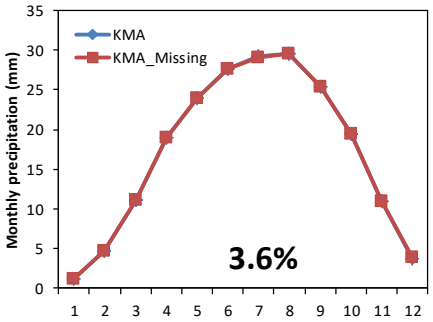


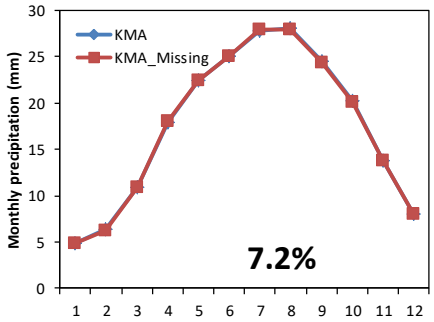


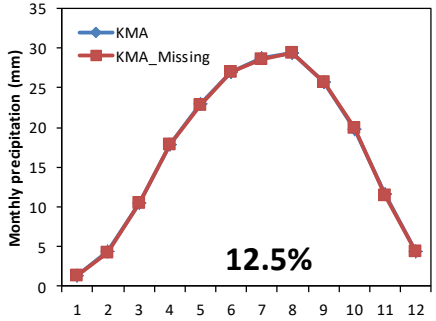
Monthly maximum temperature (30-yrs) climatology with increased missing

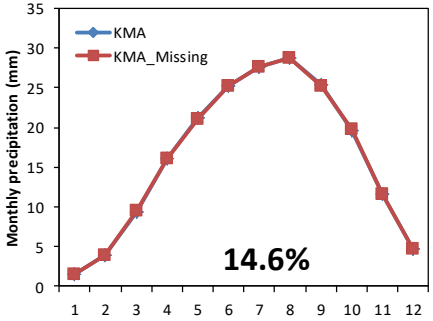


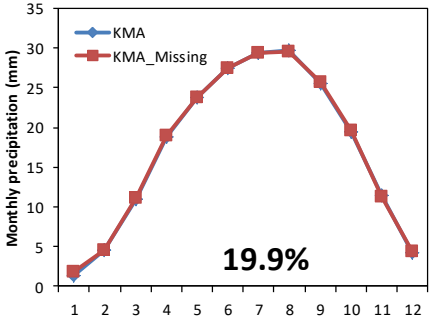


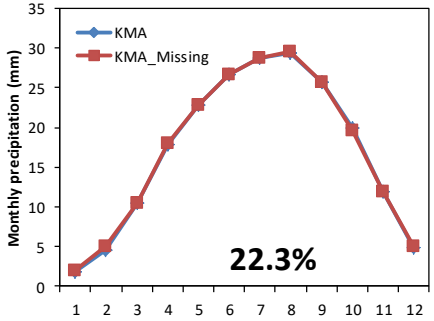


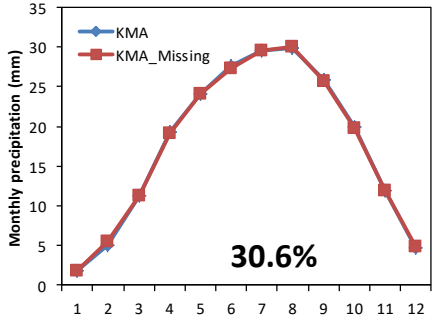


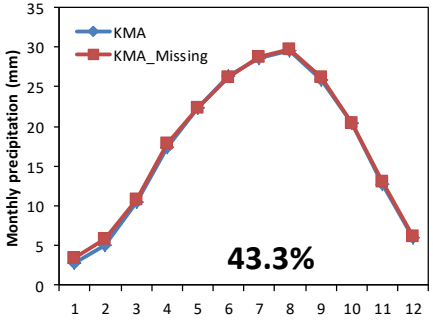




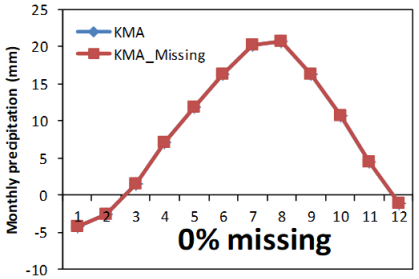


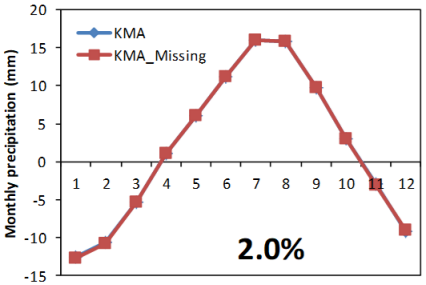


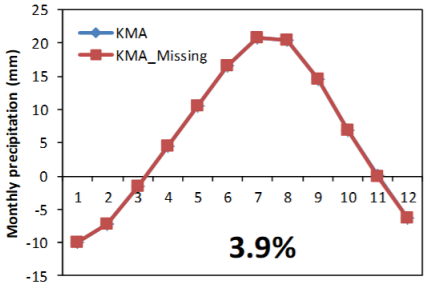


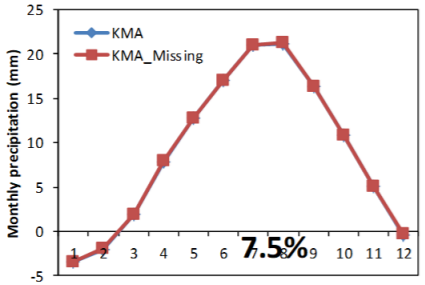


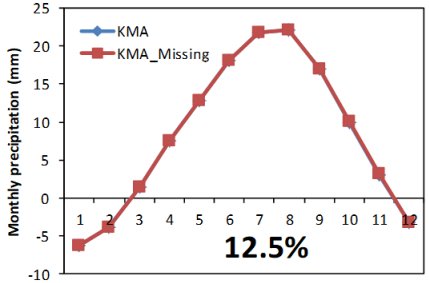
Monthly minimum temperature (30-yrs) climatology with increased missing

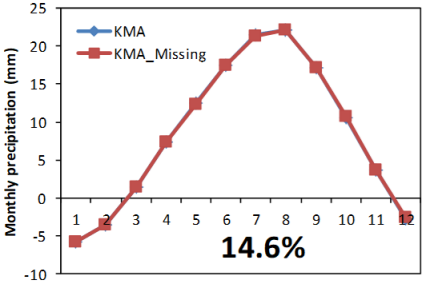


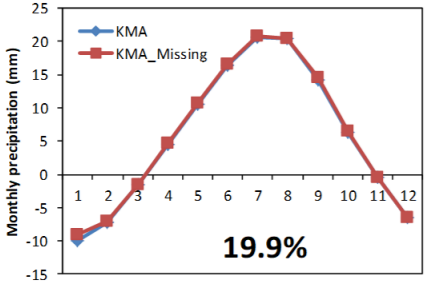


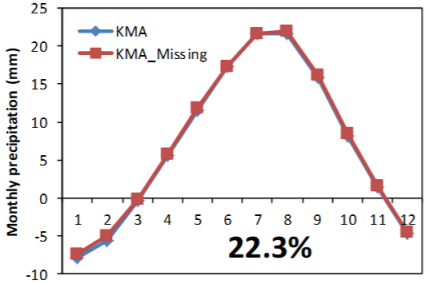


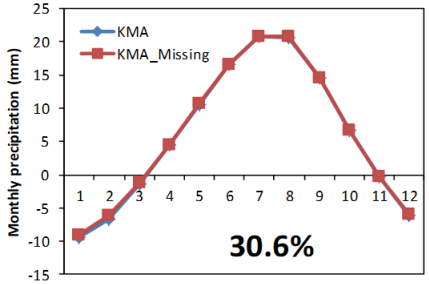


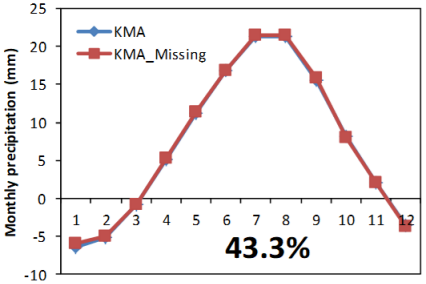






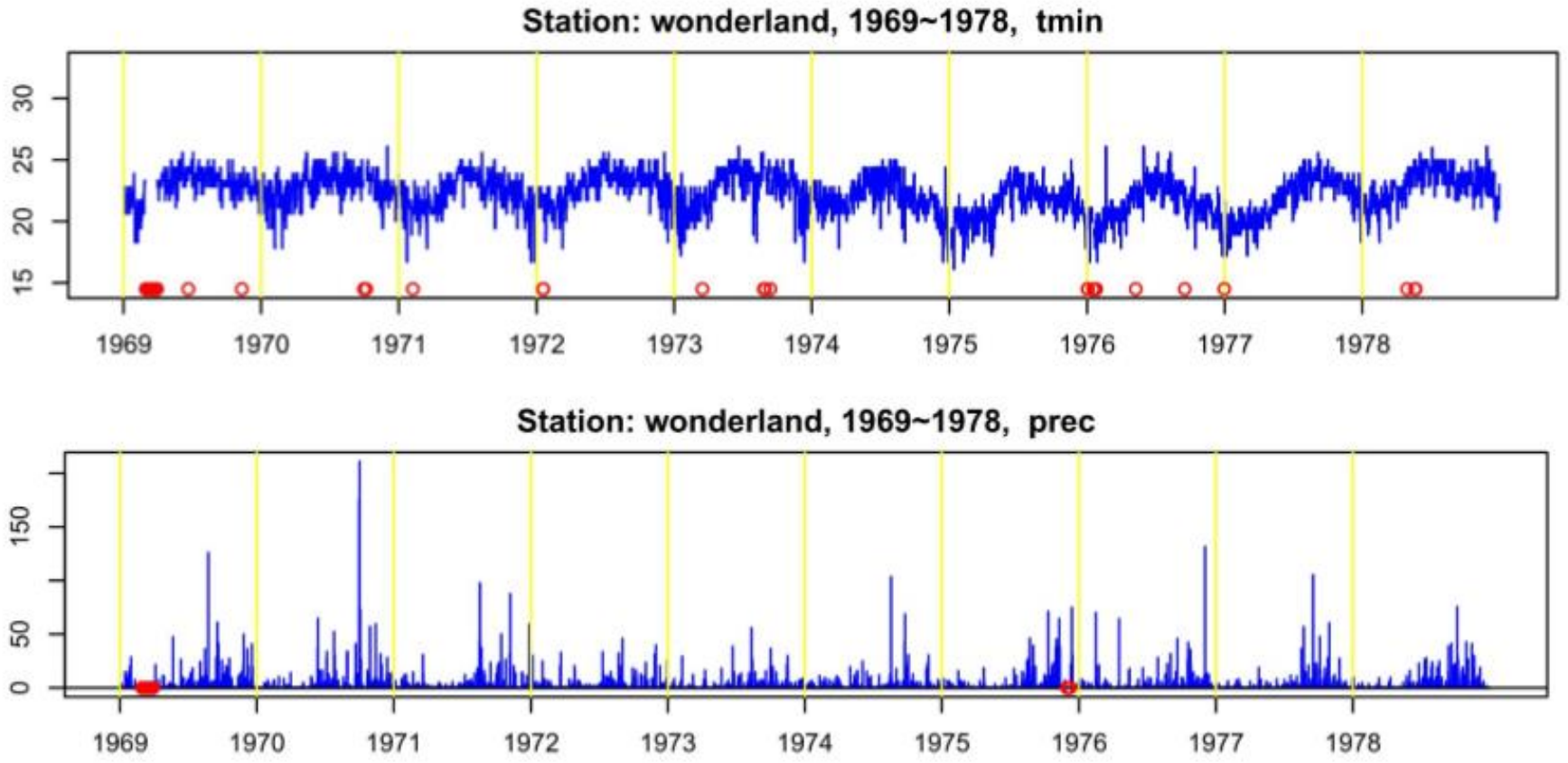






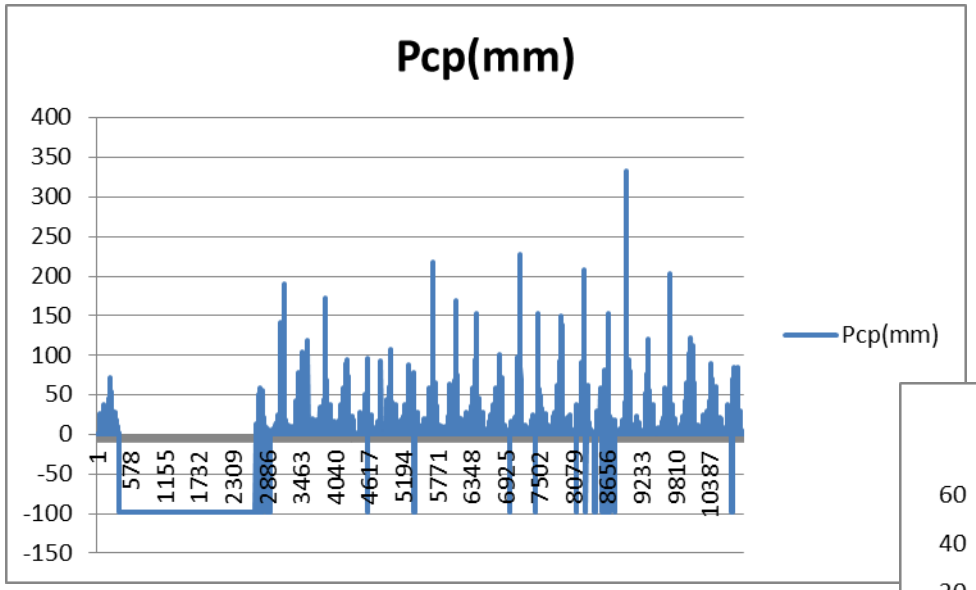
Weather station data QC algorithms

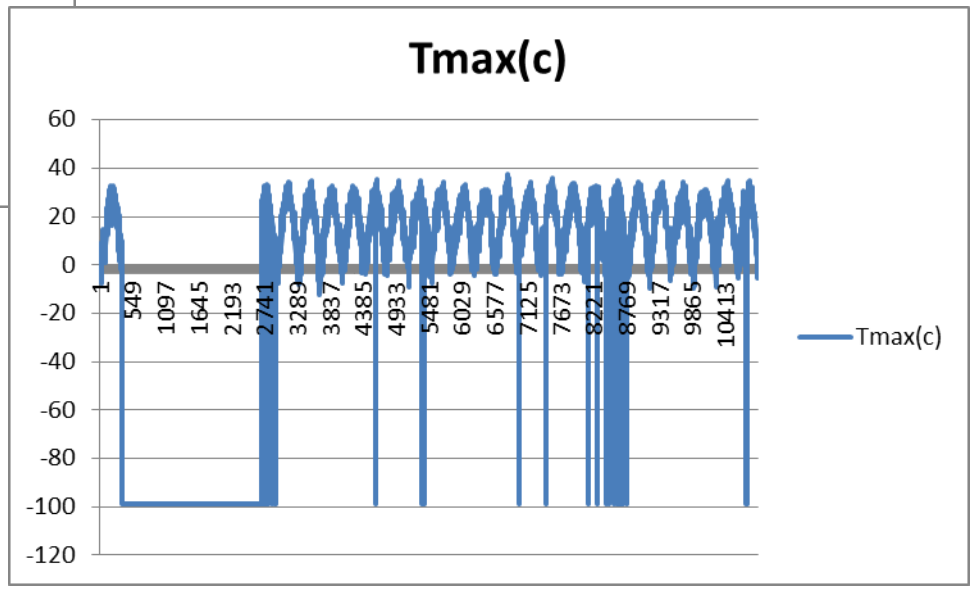
Missing



Weather station data QC algorithms

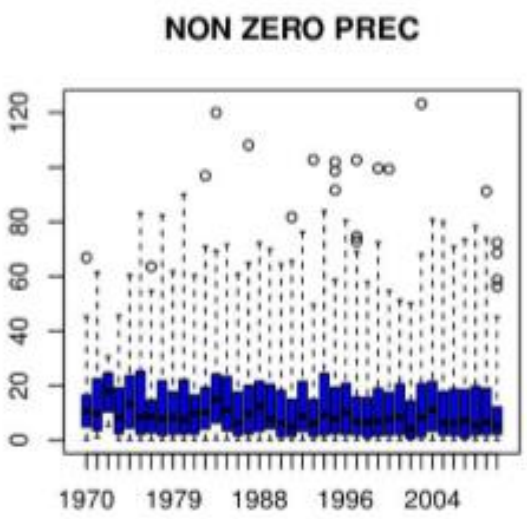
Missing

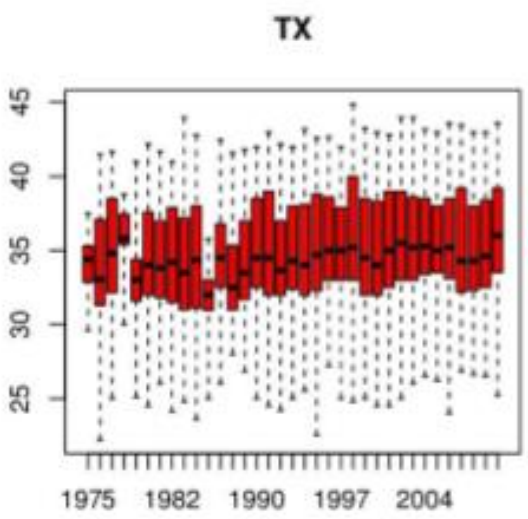


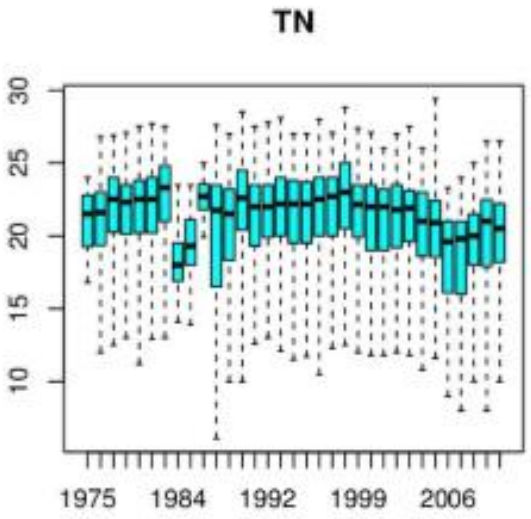


Weather station data QC algorithms

Outliers







Weather station data QC algorithms

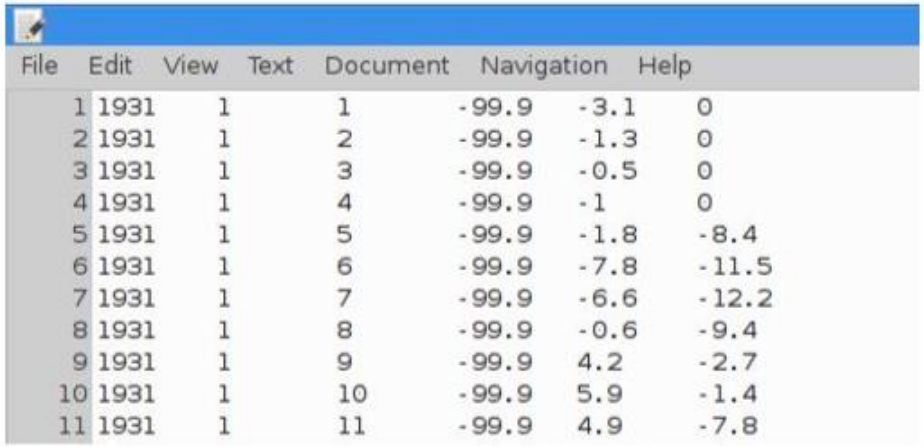
Duplicated

Identifies duplicated dates

Weather station data QC algorithms

Consistency

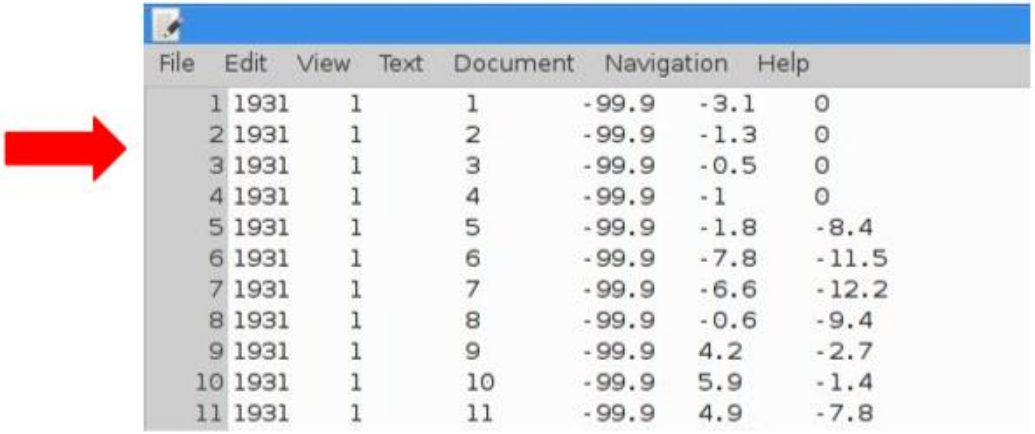
Identifies runs of 7 or more of the same value



File Edit View Text Document Navigation Help

1	1931	1	1	-99.9	-3.1	0
2	1931	1	2	-99.9	-1.3	0
3	1931	1	3	-99.9	-0.5	0
4	1931	1	4	-99.9	-1	0
5	1931	1	5	-99.9	-1.8	-8.4
6	1931	1	6	-99.9	-7.8	-11.5
7	1931	1	7	-99.9	-6.6	-12.2
8	1931	1	8	-99.9	-0.6	-9.4
9	1931	1	9	-99.9	4.2	-2.7
10	1931	1	10	-99.9	5.9	-1.4
11	1931	1	11	-99.9	4.9	-7.8

Identifies where minimum temperature > maximum temperature



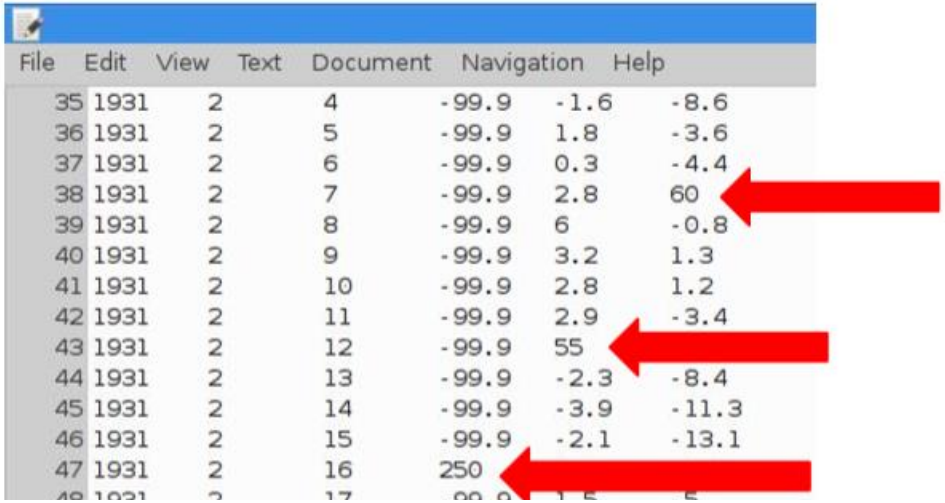
File Edit View Text Document Navigation Help

1	1931	1	1	-99.9	-3.1	0
2	1931	1	2	-99.9	-1.3	0
3	1931	1	3	-99.9	-0.5	0
4	1931	1	4	-99.9	-1	0
5	1931	1	5	-99.9	-1.8	-8.4
6	1931	1	6	-99.9	-7.8	-11.5
7	1931	1	7	-99.9	-6.6	-12.2
8	1931	1	8	-99.9	-0.6	-9.4
9	1931	1	9	-99.9	4.2	-2.7
10	1931	1	10	-99.9	5.9	-1.4
11	1931	1	11	-99.9	4.9	-7.8

Weather station data QC algorithms

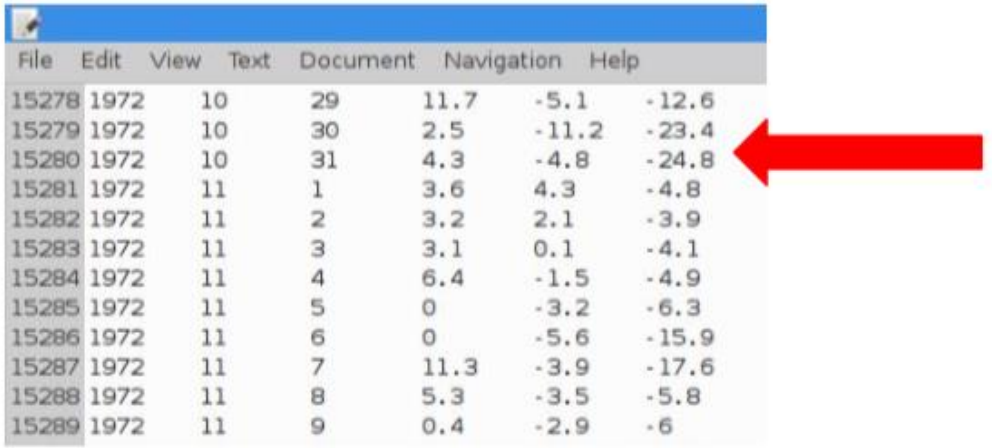
Consistency

Identifies when temperature > 50 degrees or precipitation > 200 mm



35	1931	2	4	-99.9	-1.6	-8.6
36	1931	2	5	-99.9	1.8	-3.6
37	1931	2	6	-99.9	0.3	-4.4
38	1931	2	7	-99.9	2.8	60
39	1931	2	8	-99.9	6	-0.8
40	1931	2	9	-99.9	3.2	1.3
41	1931	2	10	-99.9	2.8	1.2
42	1931	2	11	-99.9	2.9	-3.4
43	1931	2	12	-99.9	55	
44	1931	2	13	-99.9	-2.3	-8.4
45	1931	2	14	-99.9	-3.9	-11.3
46	1931	2	15	-99.9	-2.1	-13.1
47	1931	2	16	250		

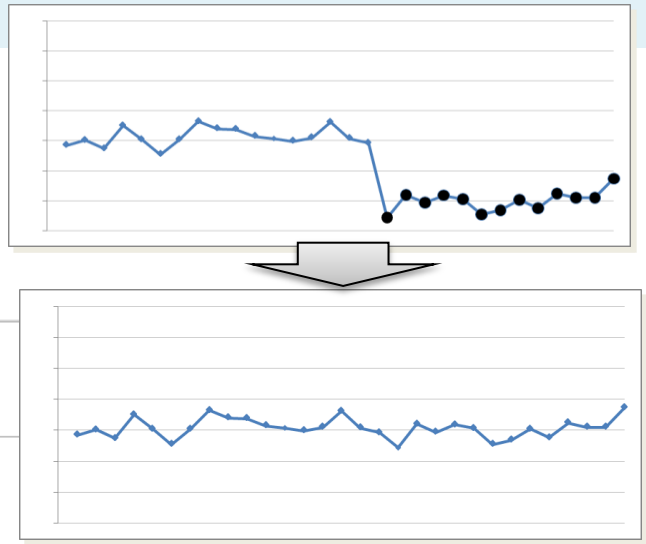
Identifies when |Δt| >= 20 degrees

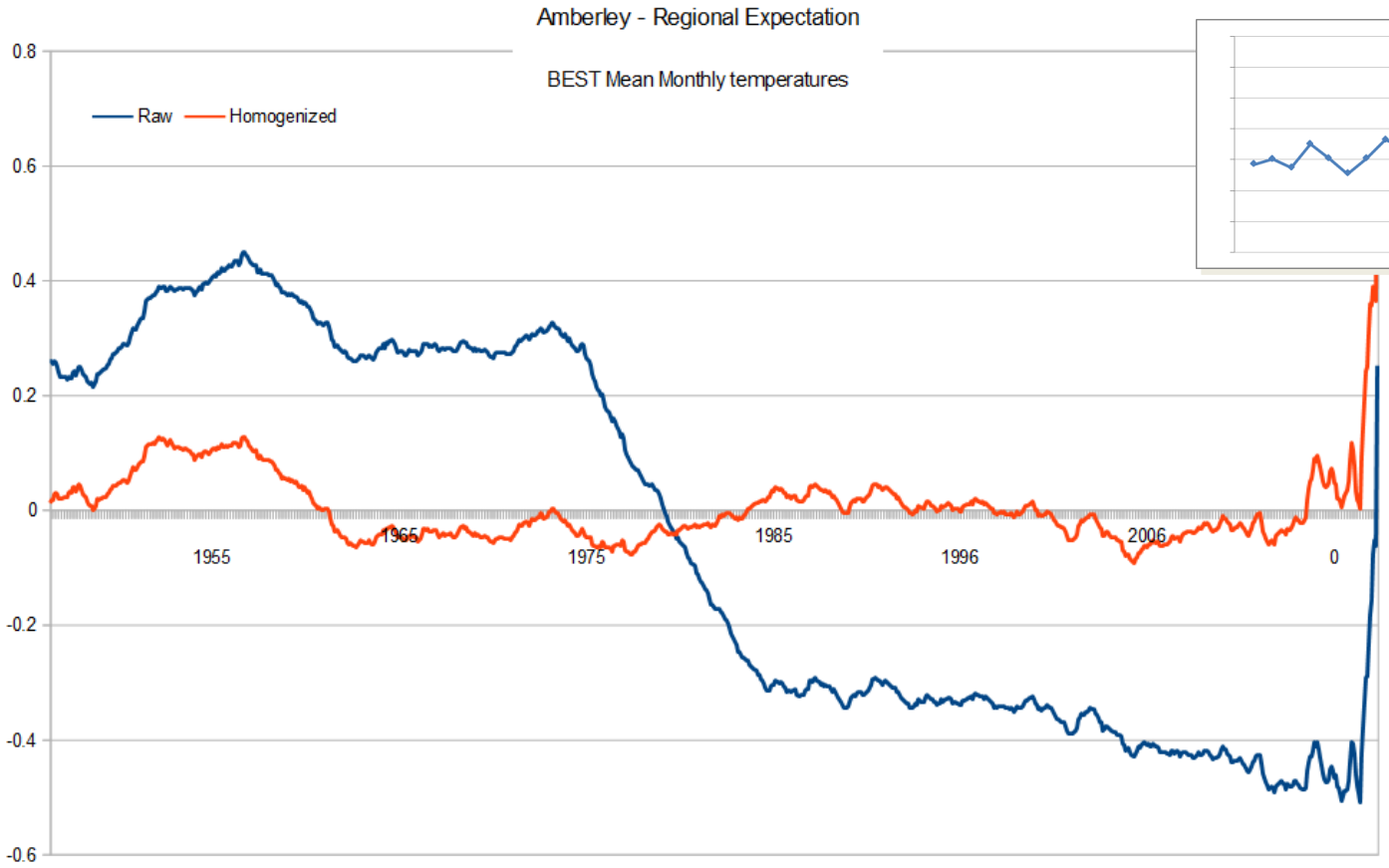


15278	1972	10	29	11.7	-5.1	-12.6
15279	1972	10	30	2.5	-11.2	-23.4
15280	1972	10	31	4.3	-4.8	-24.8
15281	1972	11	1	3.6	4.3	-4.8
15282	1972	11	2	3.2	2.1	-3.9
15283	1972	11	3	3.1	0.1	-4.1
15284	1972	11	4	6.4	-1.5	-4.9
15285	1972	11	5	0	-3.2	-6.3
15286	1972	11	6	0	-5.6	-15.9
15287	1972	11	7	11.3	-3.9	-17.6
15288	1972	11	8	5.3	-3.5	-5.8
15289	1972	11	9	0.4	-2.9	-6

Weather station data QC algorithms

Homogenization





Hands on for observed data QC

- Confirm the best correlated proxy data (in this exercise, MERRA2)
- Open Rstudio
- Install rQC package (*D:₩2018-SDTP₩Tools*)
- Open yaml and rQC_Run.R files in Rstudio
- Revise yaml for each directory path
- Run *EnvList = Set.Working.Environment* 실행
- Confirm the folder location of your observed (weather station) data: *D:₩2018-SDTP₩GridDataExt₩Observed₩User*)

Control file: yaml Configuration

Mostly default for this training, but red boxes need to be re-checked or changed

```
prjdir: D:/2018-SDTP/QualityControl
obsdir: D:/2018-SDTP/Database_Korea/Observed/User
stnfile: Station-Info.csv
bondncfile: D:/2018-SDTP/Database_Korea/grid_daily_KOR/MERRA2/MERRA2_prec_hourly_1980.nc
GirdExtractDir: D:/2018-SDTP/Database_Korea/Extracted
GirdRegionNm: KMA_ASOS_60Stns
gridnm: MERRA2
griddir: $(prjdir)/GridExtObs
QcVarNames:
  - prec          #Precipitation (mm)
  - tmax          #Max. temperature (C)
  - tmin          #Min. temperature (C)
ChkObsDir: $(prjdir)/QcOut/01_CheckObsData
ChkRgnDir: $(prjdir)/QcOut/02_CheckDataRange
ChkConsDir: $(prjdir)/QcOut/03_CheckConsistency
MinConsDay: 7
FillNeighborDir: $(prjdir)/QcOut/04_GapFillNeighbor
SrchDist_stn: 100
cc_thold_stn_pcp: 0.5
cc_thold_stn_tmp: 0.5
FillGridDir: $(prjdir)/QcOut/05_GapFillGrid
SrchDist_grid: 100
cc_thold_grid_pcp: 0.5
cc_thold_grid_tmp: 0.5
```

(1) (2) (3) (4) (5)

rQC_Run.R Script Configuration

(1) Install rQC package using Rstudio function

```
remove.packages("rGridData")

library("rQC")

# Install required packages
Install.Required.Packages()

# Setting working environment
EnvList = Set.Working.Environment (envfile='D:/2018-SDTP/Tools/QualityControl/rQC.yaml')


## Convert extracted grid data (variable-based format) into station-based format
do.call(Convert.Grid.Vars.To.Stns, EnvList)

## Check station lication based on available NC file
tryCatch(
  do.call(Check.Station.Location, EnvList),
  apcerror = function(err) {
    print(err$message)
  }
)

## Check missing data file, dates, characters
tryCatch(
  do.call(Check.Observed.Data, EnvList),
  apcerror = function(err) {
    print(err$message)
  }
```

(2)

rQC_Run.R: Extracted grid data format change (variables → stations)

yaml script

```
bondncfile: D:/2018-SDTP/Database/grid_daily_VN/MERRA2/MERRA2_prec_hourly_1980.nc
GirdExtractDir: D:/2018-SDTP/GridDataExt/Grid_Output/Extracted
GirdRegionNm: VN_Stns
gridnm: MERRA2
griddir: $(prjdir)/GridExtObs
```

rQC_Run script

```
10  ## Convert extracted grid data (variable-based format) into station-b
11  do.call(Convert.Grid.Vars.To.Stns, EnvList)
12
```

Check.Station.Location

yaml script

```
prjdir: D:/2018-SDTP/QualityControl
obsdir: D:/2018-SDTP/GridDataExt/Observed/User
stnfile: Station-Info.csv
bondncfile: D:/2018-SDTP/Database/grid_daily_VN/MERRA2/MERRA2_prec_hourly_1980.nc
```

rQC_Run script

```
13  ## Check station lication based on available NC file
14  tryCatch(
15    do.call(Check.Station.Location, EnvList),
16    apccerror = function(err) {
17      print(err$message)
18    }
19  )
```

Check.Observed.Data

yaml script

```
13  chkObsDir: $(prjdir)/QcOut/01_CheckObsData
14  chkRgnDir: $(prjdir)/QcOut/02_CheckDataRange
15  chkConsDir: $(prjdir)/QcOut/03_CheckConsistency
```

rQC_Run script

```
21  ## Check missing data file, dates, characters
22  tryCatch(
23    do.call(Check.Observed.Data, EnvList),
24    apcerror = function(err) {
25      print(err$message)
26    }
27  )
```

Check.Data.Range

yaml script

```
13  ChkObsDir: $(prjdir)/QcOut/01_CheckObsData
14  ChkRgnDir: $(prjdir)/QcOut/02_CheckDataRange
15  ChkConsDir: $(prjdir)/QcOut/03_CheckConsistency
```

rQC_Run script

```
30  ## Check data range (temperature only)
31  tryCatch(
32    do.call(Check.Data.Range, EnvList),
33    apccerror = function(err) {
34      print(err$message)
35    }
36  )
37
```

Check.Internal.Consistency

yaml script

```
13  ChkObsDir: $(prjdir)/QcOut/01_CheckObsData
14  ChkRgnDir: $(prjdir)/QcOut/02_CheckDataRange
15  ChkConsDir: $(prjdir)/QcOut/03_CheckConsistency
16  MinConsDay: 7
```

rQC_Run script

```
39  ## Internal consistency check (temperature only)
40  tryCatch(
41    do.call(Check.Internal.Consistency, EnvList),
42    apccerror = function(err) {
43      print(err$message)
44    }
45  )
46
```

Gap.Filling.Neighbor.Stations

yaml script

```yaml
QcVarNames:
  - prec            #Precipitation (mm)
  - tmax            #Max. temperature (C)
  - tmin            #Min. temperature (C)
ChkObsDir: $(prjdir)/QcOut/01_CheckObsData
ChkRgnDir: $(prjdir)/QcOut/02_CheckDataRange
ChkConsDir: $(prjdir)/QcOut/03_CheckConsistency
MinConsDay: 7
SrchDist: 100
FillNeighborDir: $(prjdir)/QcOut/04_GapFillNeighbor
r2_thold_stn: 0.2
FillGridDir: $(prjdir)/QcOut/05_GapFillGrid
r2_thold_grid: 0.1
```

rQC_Run script

```r
47  ## Gap filling using neighbor stations
48  tryCatch(
49    do.call(Gap.Filling.Neighbor.Stations, EnvList),
50    apccerror = function(err) {
51      print(err$message)
52    }
53  )
54
```

Gap.Filling.Grid.Data

yaml script

```
QcVarNames:
  - prec            #Precipitation (mm)
  - tmax            #Max. temperature (C)
  - tmin            #Min. temperature (C)
ChkObsDir: $(prjdir)/QcOut/01_CheckObsData
ChkRgnDir: $(prjdir)/QcOut/02_CheckDataRange
ChkConsDir: $(prjdir)/QcOut/03_CheckConsistency
MinConsDay: 7
SrchDist: 100
FillNeighborDir: $(prjdir)/QcOut/04_GapFillNeighbor
r2_thold_stn: 0.2
FillGridDir: $(prjdir)/QcOut/05_GapFillGrid
r2_thold_grid: 0.1
```

rQC_Run script

```
55  ## Gap filling using grid data
56  tryCatch(
57    do.call(Gap.Filling.Grid.Data, EnvList),
58    apcerror = function(err) {
59      print(err$message)
60    }
61  )
62
```

Missing rate calculation of the original observed data (before QC)

yaml script

```
prjdir: D:/2018-SDTP/QualityControl
obsdir: D:/2018-SDTP/GridDataExt/Observed/User
stnfile: Station-Info.csv
bondncfile: D:/2018-SDTP/Database/grid_daily_VN/MERRA2/MERRA2_prec_hourly_1980.nc
```

rQC_Run script

```
#### Data summary
QcVarNames <- c("prec", "tmax", "tmin")
obsdir <- "D:/2018-SDTP/Database_Korea/Observed/User"
stnfile <- "Station-Info.csv"
syear <- 1976
eyear <- 2005

# 1. Data withough QC
datadir <- "D:/2018-SDTP/QualityControl_Korea/QcOut/01_CheckObsData"
Data.Summary.Graph (obsdir, stnfile, QcVarNames, datadir, syear, eyear)

# 2. Data after QC
datadir <- "D:/2018-SDTP/QualityControl_Korea/QcOut/05_GapFillGrid"
Data.Summary.Graph (obsdir, stnfile, QcVarNames, datadir, syear, eyear)

# 3. Original grid data extracted
datadir <- "D:/2018-SDTP/QualityControl_Korea/GridExtObs"
Data.Summary.Graph (obsdir, stnfile, QcVarNames, datadir, syear, eyear)
```

Data Summary Graph of the GridExtObs data and the final data (after QC and gap filling)

rQC_Run script

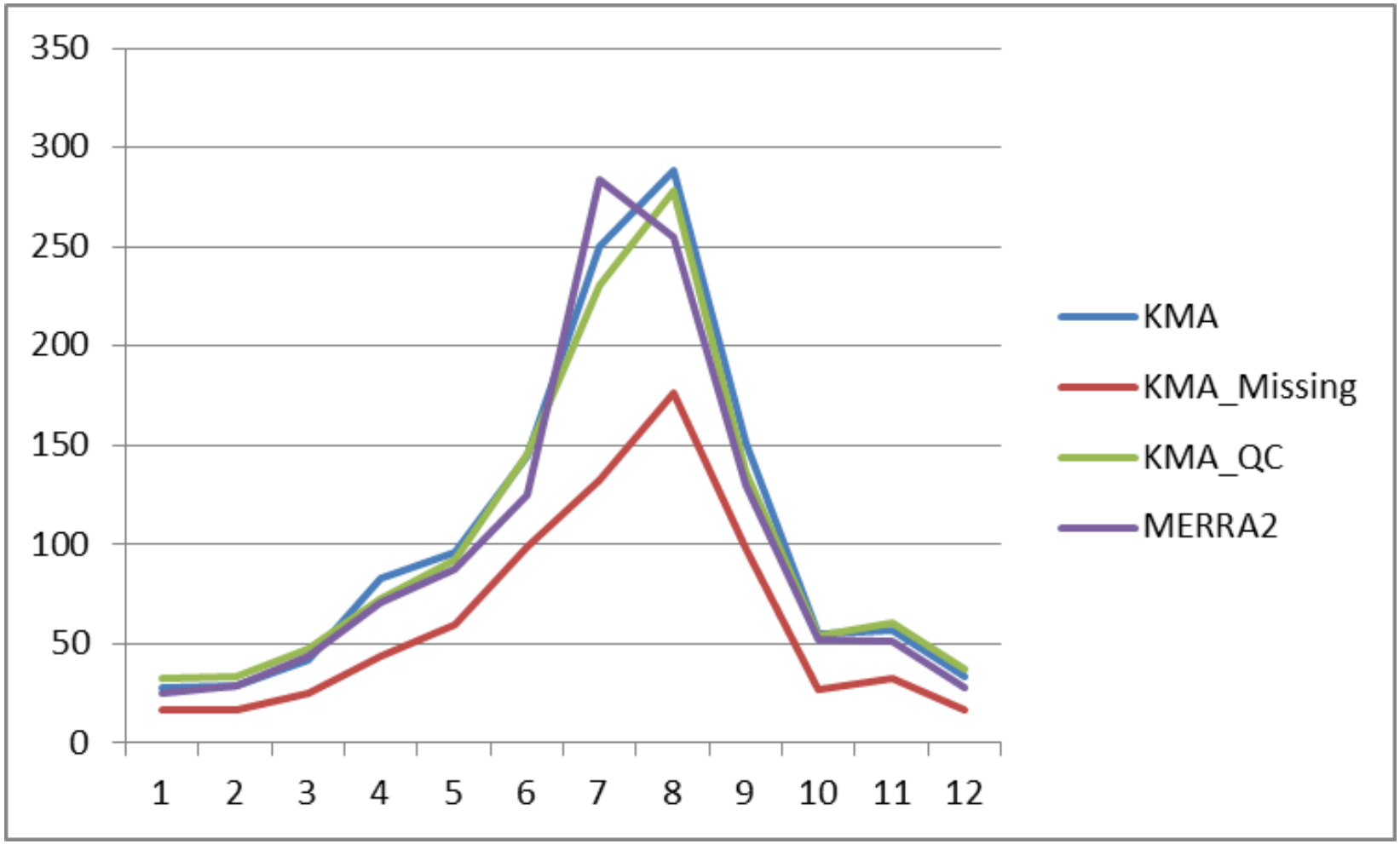
```
#### Data summary
QcVarNames <- c("prec", "tmax", "tmin")
obsdir <- "D:/2018-SDTP/Database_Korea/Observed/User"
stnfile <- "Station-Info.csv"
syear <- 1976
eyear <- 2005

# 1. Data withough QC
datadir <- "D:/2018-SDTP/QualityControl_Korea/QcOut/01_CheckObsData"
Data.Summary.Graph (obsdir, stnfile, QcVarNames, datadir, syear, eyear)

# 2. Data after QC
datadir <- "D:/2018-SDTP/QualityControl_Korea/QcOut/05_GapFillGrid"
Data.Summary.Graph (obsdir, stnfile, QcVarNames, datadir, syear, eyear)

# 3. Original grid data extracted
datadir <- "D:/2018-SDTP/QualityControl_Korea/GridExtObs"
Data.Summary.Graph (obsdir, stnfile, QcVarNames, datadir, syear, eyear)
```

QC and Gap filling result



ID129

Q&A

Any questions?

Thank You!