Matriculation No. A0169388H
Name: LI ZHI

AS.2

Q1:

1:

```
# normalize variables with min-max and z-score and decimal normalization method
#min-max
mydata<- read.table('Sediment.csv',header = T,sep = ',')
min_max<-function(x){
y<- (x-min(x,na.rm = T))/(max(x,na.rm = T)-min(x,na.rm = T))
return(y)
}
mydata_morm1<- as.data.frame(lapply(mydata,min_max))
View(mydata_morm1)
#z-score
z_score<-function(x){
y<- (x-mean(x,na.rm=T))/sd(x,na.rm=T)
return(y)
}
mydata_zscore<- as.data.frame(lapply(mydata,z_score))
#decimal
roundup<- function(x) 10^ceiling(log10(x))
decimal<- function(x){
y<- x/roundup(max(x))
return(y)
}
mydata_decimal<- as.data.frame(lapply(mydata,decimal))
norm<-summary(mydata_morm1)
write.table(norm,file = 'summary of norm.txt')
deci<-summary(mydata_decimal)
write.table(deci,file = 'summary of decimal.txt')
zso<-summary(mydata_zscore)
write.table(zso,file = 'summary of z_score.txt')
```

Tab.1 summary of min-max normalization

	theta	theta_p	nu	uf	uf_p	ws	d50	cb	bnch
Min	0	0	0	0	0	0	0	0	0
1st Qu.	0.08538	0.08635	0.1328	0.1764	0.1719	0.1465	0.1371	0.02202	0.04433
Median	0.13654	0.18951	0.1577	0.2747	0.3419	0.2514	0.3053	0.10870	0.15973
Mean	0.17744	0.29229	0.1917	0.3024	0.3913	0.3087	0.3163	0.18855	0.30247
3rd.Qu	0.21391	0.45907	0.2116	0.4178	0.6009	0.4144	0.3998	0.31789	0.53107
Max.	1	1	1	1	1	1	1	1	1

Tab.2 summary of z-score normalization

	theta	theta_p	nu	uf	uf_p	ws	d50	cb	bnch
Min	-1.2082	-1.1678	-1.4461	-1.7212	-1.6061	-1.5794	-1.3175	-0.9210	-1.0161
1st Qu.	-0.6268	-0.8228	-0.4443	-0.7173	-0.9006	-0.8297	-0.7465 3	-0.8135	-0.8672

Median	-0.2785	-0.4106	-0.2563	-0.1575	-0.2025	-0.2933	-0.04602	-0.3901	-0.4795
Mean	0	0	0	0	0	0	0	0	0
3^{rd}.Qu	0.2484	0.6663	0.1504	0.6567	0.8603	0.5407	0.34742	0.6318	0.7680
Max.	5.6011	2.8275	6.0988	3.9704	2.4987	3.5364	2.84719	3.9639	2.3433

Tab.3 summary of decimal normalization

	theta	theta_p	nu	uf	uf_p	ws	d50	cb	bnch
Min	0.00614	0.00463	0.07339	0.02907	0.01802	0.1451	0.01710	0.00014	0.000003
1^{st} Qu.	0.05751	0.02133	0.10312	0.05886	0.03545	0.2481	0.02900	0.01210	0.014186
Median	0.08829	0.04128	0.10870	0.07547	0.05270	0.3218	0.04360	0.05918	0.051109
Mean	0.11289	0.06116	0.11631	0.08014	0.05770	0.3621	0.04456	0.10255	0.096782
3^{rd}.Qu	0.13484	0.09341	0.12077	0.09963	0.07896	0.4364	0.05180	0.17280	0.169922
Max.	0.60779	0.19802	0.29729	0.19796	0.11944	0.8480	0.10390	0.54329	0.319959

2.

```
library(ggplot2)
data_com<-as.data.frame(c(mydata_decimal,mydata_morm1))
ggplot(data_com,aes(theta,theta.1))+geom_point()+ggtitle('scatter plot')
```

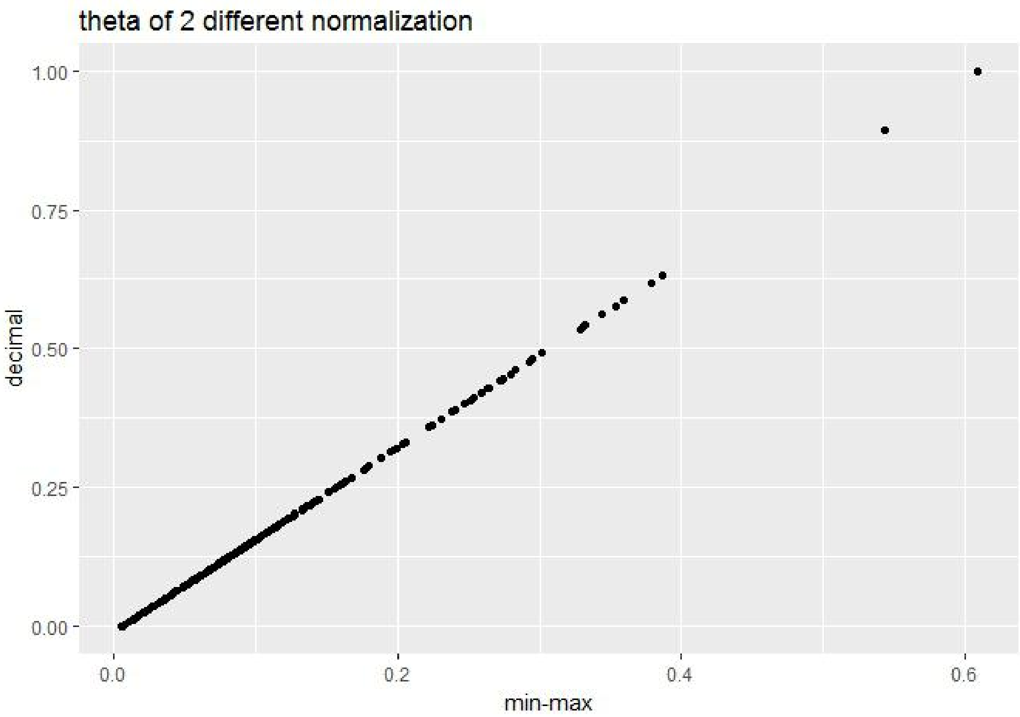

Fig.1

3.

```
library(reshape2)
long_mydata<-melt(mydata_morm1)
str(long_mydata)
ggplot(long_mydata,aes(x=variable,y=value))+geom_boxplot()+ggtitle('all variables of nomalization')
```

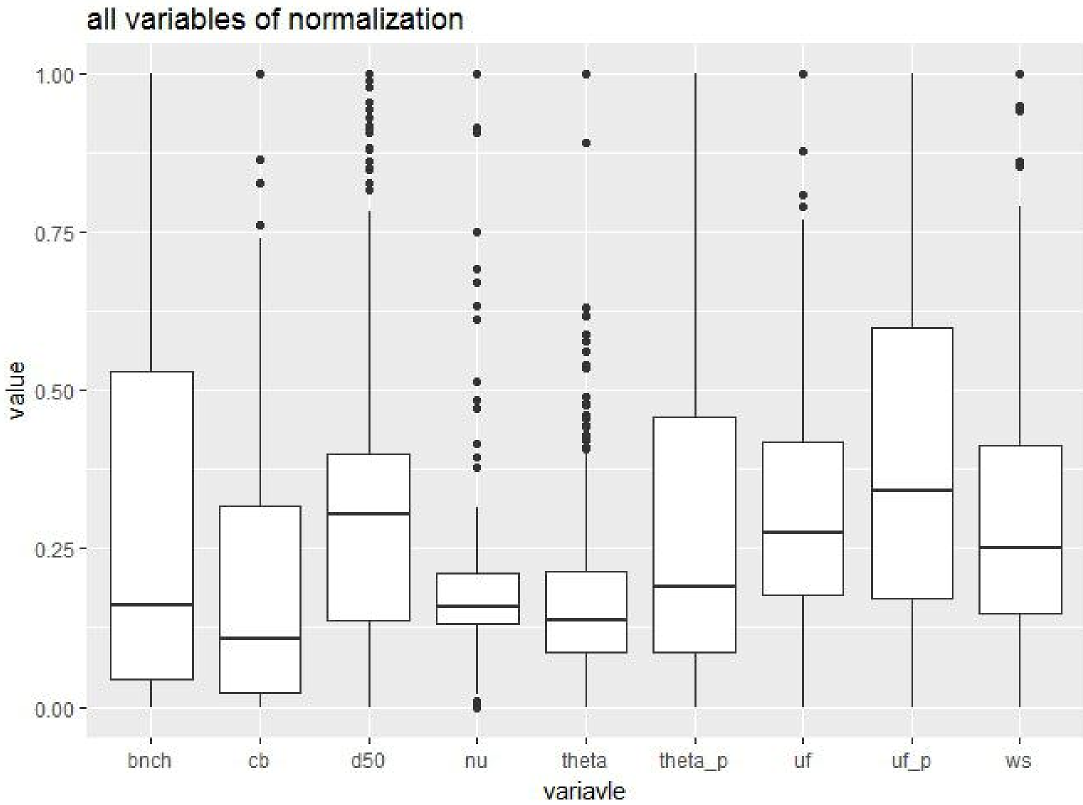

Fig.2

```
##list all outliers
outliers<- function(x){
   y<- boxplot.stats(x)$out
   return(y)
}
out<-lapply(mydata_morm1,outliers)
```

Tab.3 outliers of all variables

x	theta	nu	uf	ws	D50	cb
1	0.428737638161722	0.394185719326297	0.809935460950915	0.863138426518708	0.955069124423963	0.8660406885759
2	0.631962104213413	0.377615600030371	1	0.854460093896714	0.817972350230415	0.762312436711774
3	0.491731072882905	0.473016610316351	0.792113209781515	0.941812491108266	0.862903225806452	0.827782380557857
4	0.410388099393335	0.672215349022095	0.87832316892652	0.950917626973965	0.908986175115207	1
5	0.446605169118258	0.908708021992256		1	1	

6	0.587750353195379	0			0.919354838709677	
7	0.430333250228538	0.751046240012148			0.98963133640553	
8	0.461015540596692	1			0.852534562211982	
9	0.539034322280396	0.473016610316351			0.944700460829493	
10	0.411352115017037	0.614108273000532			0.884792626728111	
11	0.420094739466467	0.917015413337383			0.862903225806452	
12	0.618681957948974	0.51450890364766			0.933179723502304	
13	0.40701404471038	0.377615600030371			0.915898617511521	
14	0.476323443862711	0.63487675136335			0.849078341013825	
15	0.577395495720103	0.414954197689116			0.908986175115207	
16	0.480744618964514	0.00829845866626174			0.884792626728111	
17	0.441668744286545	0.00829845866626174			0.955069124423963	
18	0.454932269591956	0.00414922933313087			0.862903225806452	
19	0.892595362752431	0.614108273000532			0.862903225806452	
20	0.427973074046372	0.692939163990585			0.884792626728111	
21	0.444195130059004	0.485477697334042			0.828341013824885	
22	0.562137455331173				0.97926267281106	
23	0.445026178010471				0.862903225806452	
24	0.535593783761323				0.881336405529954	
25	1.000000000000000					
26	0.54167705476606					

No outliers in theta_p.

4.

```
ggplot(mydata_morm1,aes(theta,theta_p))+geom_point()+geom_smooth(method =
'loess')+ggtitle('theta versus theta_p in nomalization')+xlab('theta')+ylab('theta_p')
```

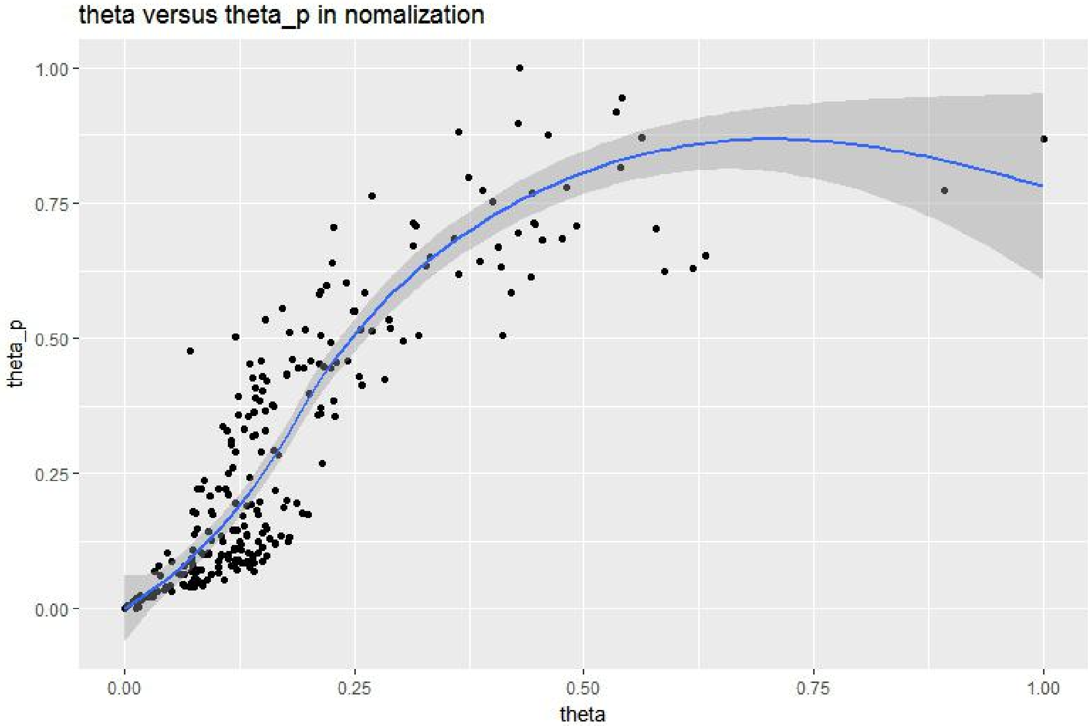

Fig.3

Q2:

1.

```
mydata2<-read.csv('Global_monthly_mean_temperature_anomaly.csv')
attach(mydata2)
GCAG_data<-mydata2[Source=='GCAG',]
GIS_data<-mydata2[Source=='GISTEMP',]
library(zoo)
GCAG_data_new<-GCAG_data[order(GCAG_data$Date),]
GCAG_data_Jan<-
subset(GCAG_data_new,format(as.yearmon(GCAG_data_new$Date),"%b")=='1 月')
GIS_data_Jan<-subset(GIS_data,format(as.yearmon(GIS_data$Date),"%b")=='1 月')
new_data<-rbind(GIS_data_Jan,GCAG_data_Jan)
ggplot(new_data,aes(x=as.yearmon(new_data$Date),y=Mean))+
  scale_x_yearmon(breaks=waiver(),labels=date_format("%b %Y"))+
  geom_line(aes(color=Source,group=Source))+xlab('Date')+ylab('Mean')+ggtitle('Time series
of different source in Jan')
```

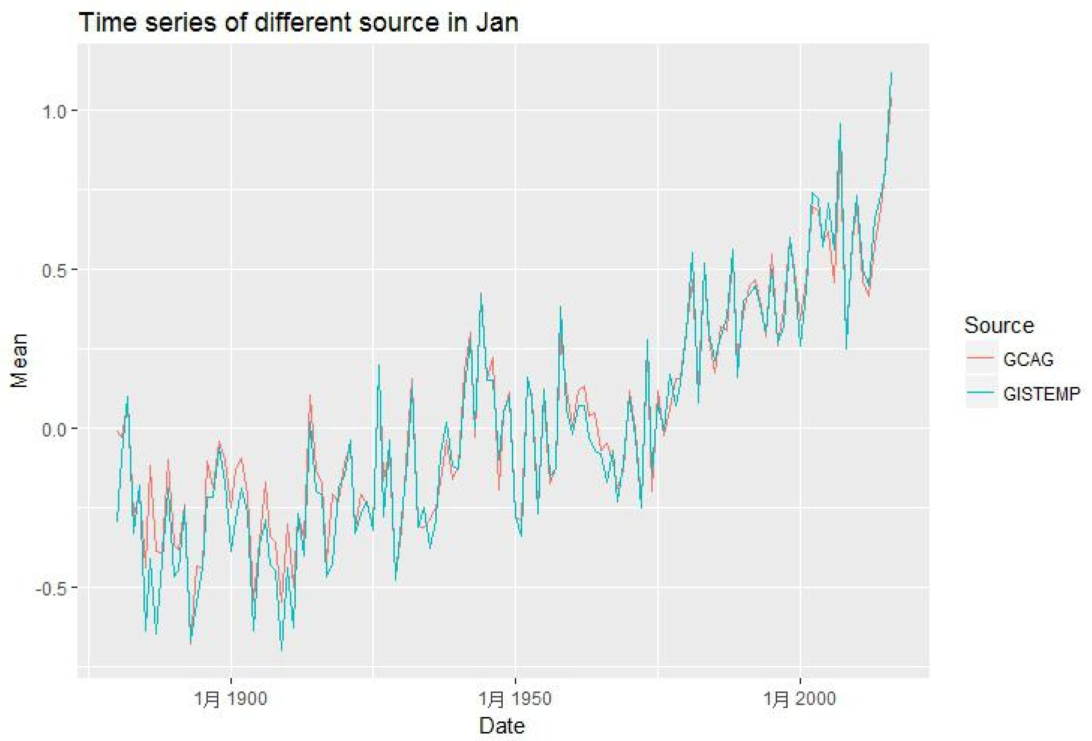

Fig.4

```
##detrend
GCAG_detrend<- as.list(diff(GCAG_data_Jan$Mean),colnames('Mean'))
GIS_detrend<- as.list(diff(GIS_data_Jan$Mean),colnames('Mean'))
detrend<- rbind(GCAG_detrend,GIS_detrend)
GIS_data_Jan$Mean[2:137]<- GIS_detrend
GIS_new<-GIS_data_Jan[2:137,]
GCAG_data_Jan$Mean[2:137]<-GCAG_detrend
GCAG_new<-GCAG_data_Jan[2:137,]
last_data<-rbind(GCAG_new,GIS_new)
last_data$Mean<- as.numeric(last_data$Mean)
ggplot(last_data,aes(x=as.yearmon(last_data$Date),y=Mean))+
  scale_x_yearmon(breaks=waiver(),labels=date_format("%b %Y"))+
   geom_line(aes(color=Source,group=Source))+xlab('Date')+ylab('Mean')+ggtitle('Time  series
of different source in Jan after detrending')
```

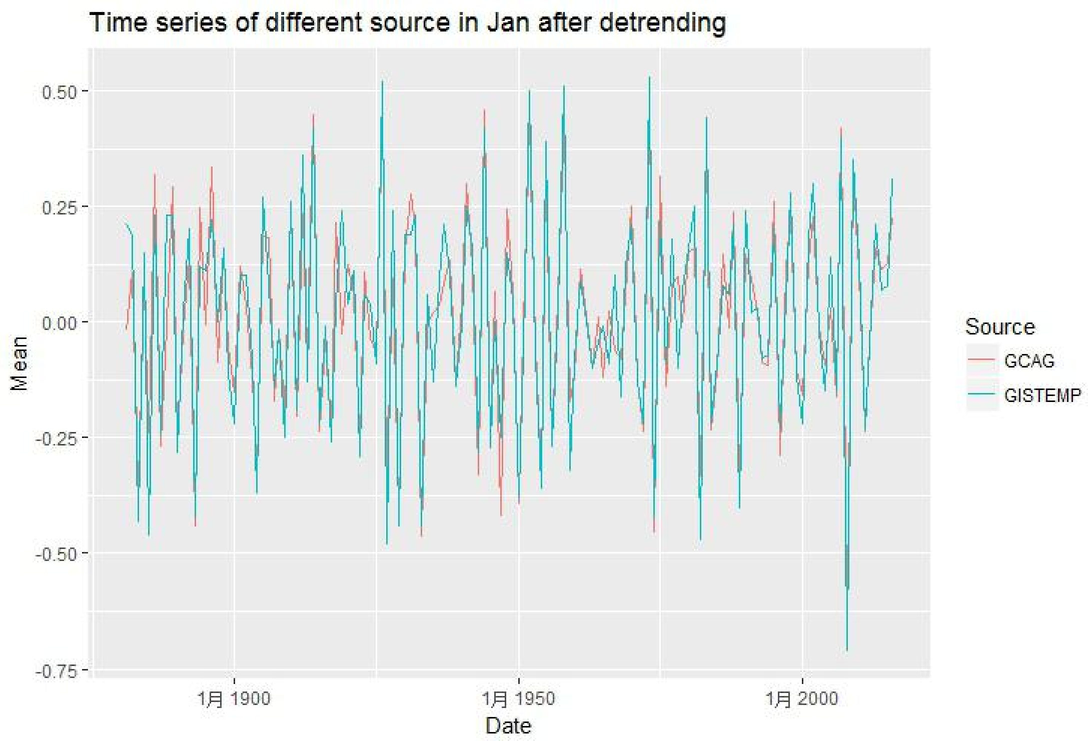

Fig.5

2.

```
### delete 2016
Year_1<-format(as.yearmon(GCAG_data_new$Date),'%Y')
Year_1<- as.numeric(Year_1)
GCAG_data_new$Date<-Year_1
GCAG_data_new<-GCAG_data_new[GCAG_data_new$Date<2016,]
#calculate anual temperature.
GCAG_data$Date<-as.yearmon(GCAG_data$Date)
GIS_data$Date<-as.yearmon(GIS_data$Date)
mean_anual_GCAG<-vector('numeric',137)
mean_anual_GIS<-vector('numeric',137)
  for(i in 1:137){
    GCAG_s<-GCAG_data[year(GCAG_data$Date)==1879+i,]
    GIS_s<-GIS_data[year(GIS_data$Date)==1879+i,]
  mean_anual_GCAG[i]<-mean(GCAG_s$Mean)
  mean_anual_GIS[i]<-mean(GIS_s$Mean)
  }
mean_anual_GCAG<-as.data.frame(mean_anual_GCAG,col.names='mean')
mean_anual_GIS<-as.data.frame(mean_anual_GIS,col.names='mean')
GIS<-rep('GIS',length(mean_anual_GCAG$mean))
GCAG<-rep('GCAG',length(mean_anual_GCAG$mean))
colnames(mean_anual_GCAG)<-c('mean','Source')
colnames(mean_anual_GIS)<-c('mean','Source')
mean_anual_GCAG$Source<-GCAG
mean_anual_GIS$Source<-GIS
```

```
#plot time series in one figure
library(reshape2)
library(ggplot2)
mean_GCAG_GIS<-rbind(mean_anual_GIS,mean_anual_GCAG)
new_data$Mean<-mean_GCAG_GIS$mean
ggplot(new_data,aes(x=Date,y=Mean))+geom_line(aes(color=Source,group=Source))+
  scale_x_yearmon(breaks=waiver(),labels=date_format("%Y"))+ggtitle('Time series of anual mean')+
```

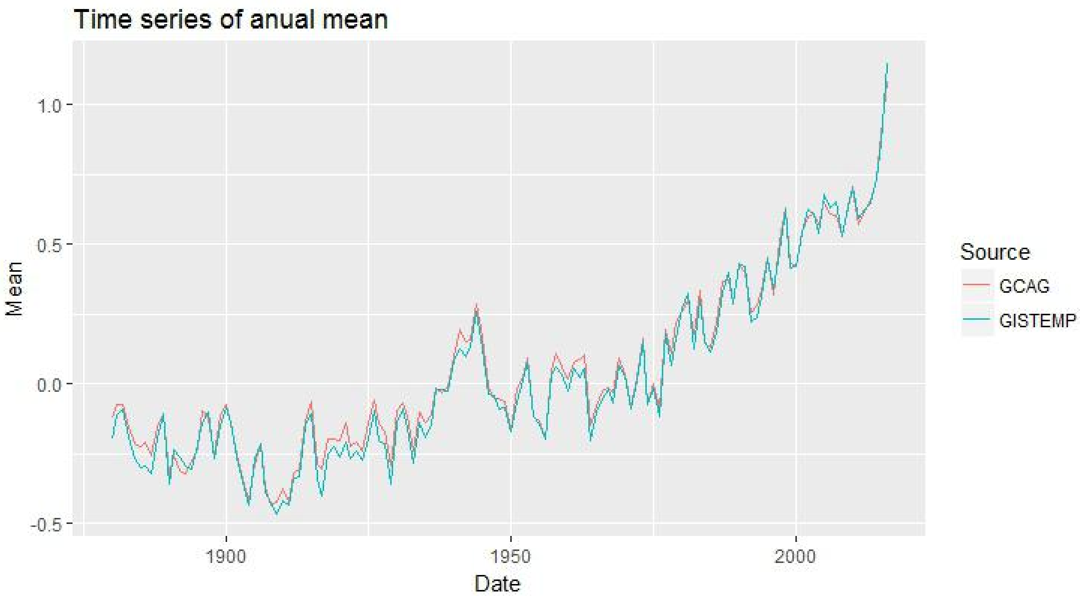

Fig.6

```
mean_anual_GCAG_detrend<- as.data.frame(diff(mean_anual_GCAG$mean_anual_GCAG))
mean_anual_GIS_detrend<-as.data.frame(diff(mean_anual_GIS$mean_anual_GIS))
mean_anual_GCAG_detrend$Date<-GCAG_data_Jan$Date[2:137]
mean_anual_GCAG_detrend$Source<-GCAG_data_Jan$Source[2:137]
colnames(mean_anual_GCAG_detrend)<-c('mean','date','source')
mean_anual_GIS_detrend$Date<-GCAG_data_Jan$Date[2:137]
mean_anual_GIS_detrend$Source<-GIS_data$Source[1:136]
colnames(mean_anual_GIS_detrend)<- c('mean','date','source')
mean_anual_all<-rbind(mean_anual_GIS_detrend,mean_anual_GCAG_detrend)
ggplot(mean_anual_all,aes(x=date,y=mean))+geom_line(aes(color=source,group=source))+
scale_x_yearmon(breaks=waiver())+ggtitle('Time series of anual mean after detrending')
```

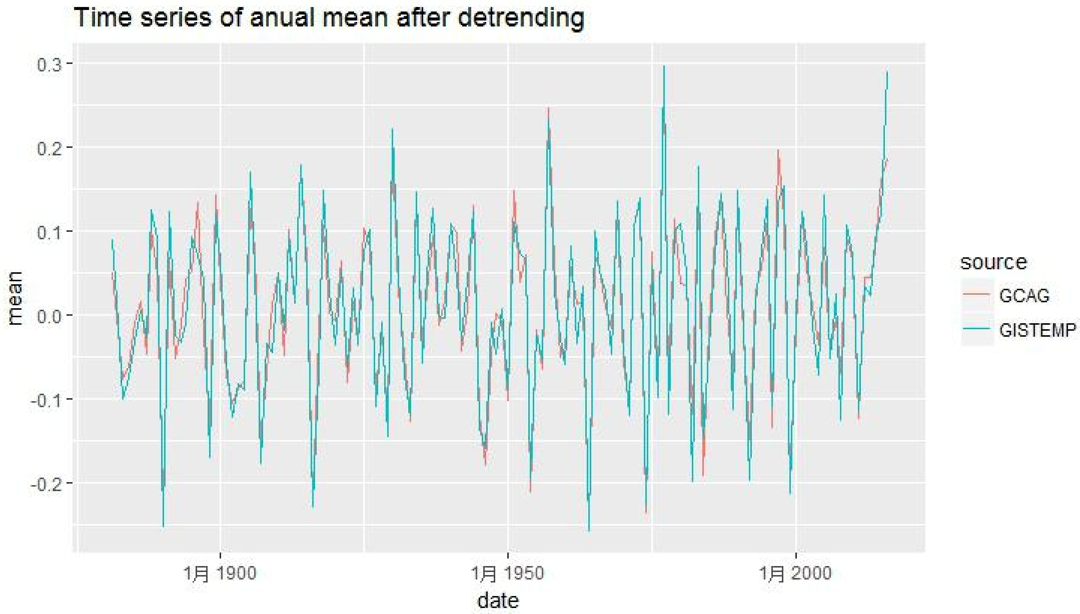
Time series of anual mean after detrending
0.3
0.2
0.1
0.0
-0.1
-0.2
mean
1月 1900
1月 1950
1月 2000
date
source
GCAG
GISTEMP

Fig.7