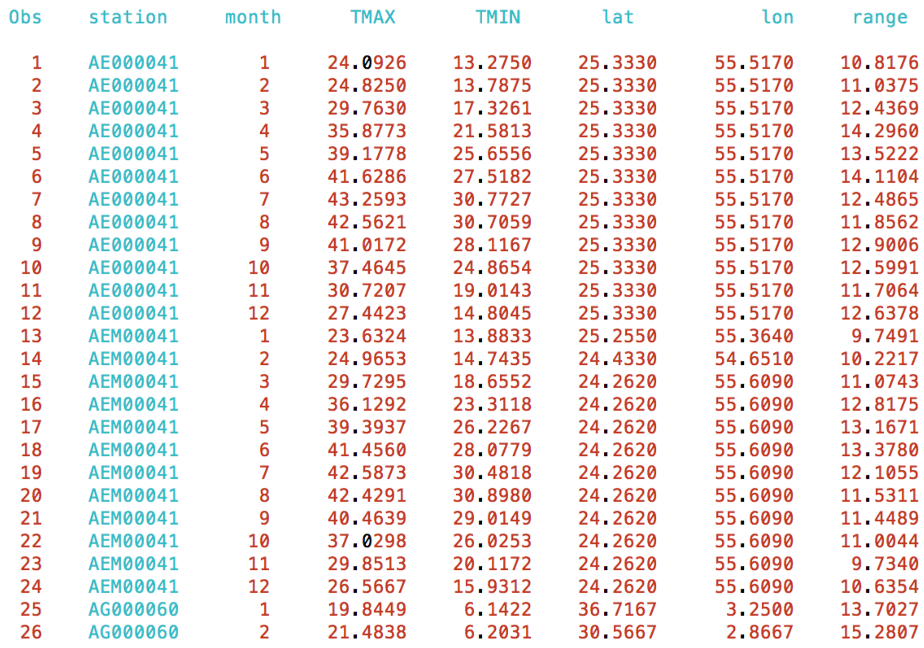
Problem1:
step1: preparation

code:

```
libname mydata '~/sas/';
filename ghcnd_gz pipe "gzip -dc 2014.csv.gz" lrecl=80;

data ghcnd;
infile ghcnd_gz delimiter=",";
input station $ date : yymmdd8. obstype $ obsval;
format date mmddyy10.;
month = month(date);
if obstype = "TMAX" or obstype = "TMIN";
obsval = obsval/10;

data stations;
infile "ghcnd-stations.txt";
input station $ 1-11 lat 13-20 lon 22-30 elev 32-37 state $ 39-40;

proc sort data=ghcnd out=ghcnd2;
by station;

proc sort data=stations out=stations2;
by station;

data ghcnd3;
merge ghcnd2(in=x) stations2(in=y);
by station;
if x=1 and y=1;

proc summary data=ghcnd3 nway;
class station month obstype;
output out=ghcnd4
       mean(obsval)= meanobsval;

proc transpose data=ghcnd4(drop=_TYPE_ _FREQ_) out=ghcnd5;
by station month;
id obstype;
var meanobsval;

data ghcnd6(drop=_NAME_ elev state);
merge ghcnd5(in=x) stations2(in=y);
by station;
if x=1 and y=1;
range=TMAX-TMIN;


data mydata.result1;
set ghcnd6;
```

```
proc print data=ghcnd6;
```

save the obtained dataset as result1.sas7bdat for later use, not to calculate every time, since this step is time-consuming, i also print some of the result of the obtained dataset in the following picture:



```
Obs   station    month    TMAX      TMIN      lat        lon      range

  1   AE000041     1     24.0926   13.2750   25.3330   55.5170   10.8176
  2   AE000041     2     24.8250   13.7875   25.3330   55.5170   11.0375
  3   AE000041     3     29.7630   17.3261   25.3330   55.5170   12.4369
  4   AE000041     4     35.8773   21.5813   25.3330   55.5170   14.2960
  5   AE000041     5     39.1778   25.6556   25.3330   55.5170   13.5222
  6   AE000041     6     41.6286   27.5182   25.3330   55.5170   14.1104
  7   AE000041     7     43.2593   30.7727   25.3330   55.5170   12.4865
  8   AE000041     8     42.5621   30.7059   25.3330   55.5170   11.8562
  9   AE000041     9     41.0172   28.1167   25.3330   55.5170   12.9006
 10   AE000041    10     37.4645   24.8654   25.3330   55.5170   12.5991
 11   AE000041    11     30.7207   19.0143   25.3330   55.5170   11.7064
 12   AE000041    12     27.4423   14.8045   25.3330   55.5170   12.6378
 13   AEM00041     1     23.6324   13.8833   25.2550   55.3640    9.7491
 14   AEM00041     2     24.9653   14.7435   24.4330   54.6510   10.2217
 15   AEM00041     3     29.7295   18.6552   24.2620   55.6090   11.0743
 16   AEM00041     4     36.1292   23.3118   24.2620   55.6090   12.8175
 17   AEM00041     5     39.3937   26.2267   24.2620   55.6090   13.1671
 18   AEM00041     6     41.4560   28.0779   24.2620   55.6090   13.3780
 19   AEM00041     7     42.5873   30.4818   24.2620   55.6090   12.1055
 20   AEM00041     8     42.4291   30.8980   24.2620   55.6090   11.5311
 21   AEM00041     9     40.4639   29.0149   24.2620   55.6090   11.4489
 22   AEM00041    10     37.0298   26.0253   24.2620   55.6090   11.0044
 23   AEM00041    11     29.8513   20.1172   24.2620   55.6090    9.7340
 24   AEM00041    12     26.5667   15.9312   24.2620   55.6090   10.6354
 25   AG000060     1     19.8449    6.1422   36.7167    3.2500   13.7027
 26   AG000060     2     21.4838    6.2031   30.5667    2.8667   15.2807
```

(a)

code:

```
libname mydata '~/sas/';

data data1;
set mydata.result1;

proc summary data=data1;
class station;
output out=data2
       max(TMAX)=maxtmax
       min(TMAX)=mintmax
       max(TMIN)=maxtmin
       min(TMIN)=mintmin;


data data3(drop=_TYPE_ _FREQ_);
  set data2;
  rangemax=maxtmax-mintmax;
  rangemin=maxtmin-mintmin;

data data4;
set data3;
difference=rangemax-rangemin;

proc summary data=data4;
output out=data5(drop=_TYPE_ _FREQ_)
       maxid(difference(station))=station
       max(difference)=maxrange;

proc summary data=data4;
output out=data6(drop=_TYPE_ _FREQ_)
       minid(difference(station))=station
       min(difference)=minrange;

proc print data=data5;

proc print data=data6;

proc print data=data4;
```

by the code above, i get the following result:

```
                                   Obs    station     maxrange

                                     1    CA006059     32.0495
                                            The SAS System                        19:41 Sunday, November 29, 2015   2

                                   Obs    station     minrange

                                     1    USS0006H    -266.395
                                            The SAS System                        19:41 Sunday, November 29, 2015   3
Obs    station     maxtmax     mintmax    maxtmin     mintmin    rangemax    rangemin    difference

  1                46.6741    -99.9000    281.101    -99.9000     146.574     381.001      -234.427
  2    AE000041    43.2593     24.0926     30.773     13.2750      19.167      17.498         1.669
  3    AEM00041    42.5873     23.6324     30.898     13.8833      18.955      17.015         1.940
  4    AG000060    38.0852     17.9635     24.836      4.8167      20.122      20.019         0.103
  5    AGE00147    35.1144     17.1175     23.907      8.1792      17.997      15.728         2.269
  6    AGM00060    37.4329     14.7476     22.231      4.3783      22.685      17.852         4.833
  7    AJ000037    34.4471      6.9818     22.176      2.0921      27.465      20.083         7.382
  8    ALM00013    31.6120     15.1762     17.705      1.7650      16.436      15.940         0.496
  9    AM000037    27.0194     -3.1966     13.727    -12.2913      30.216      26.018         4.198
 10    AMM00037    24.5167     -0.6621     15.238     -7.2789      25.179      22.516         2.662
 11    AO000066    31.9000     24.7900     24.000     15.3000       7.110       8.700        -1.590
 12    AQW00061    31.2500     28.5839     25.958     24.0516       2.666       1.906         0.760
 13    AR000000    35.2263     19.3545     22.090      9.5368      15.872      12.553         3.319
 14    AR000087    30.7721     16.5293     16.336      2.6523      14.243      13.684         0.559
 15    AR000870    31.0500     18.2667     16.804      2.3148      12.783      14.489        -1.706
 16    AR000875    31.7692     19.6167     21.200      7.7500      12.153      13.450        -1.297
 17    AR000877    34.9235     15.5111     16.804      2.5733      19.412      14.230         5.182
```

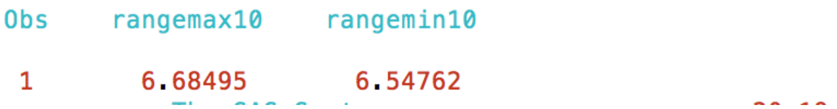
so we can see that CA006059 have the largest range, meanwhile USS0006H have the least range.

(b)

insert the following code the part(a) code:

```
proc univariate data=data3;
var rangemax rangemin;
output out=data7 pctlpts=10 pctlpre=rangemax rangemin;

proc print data=data7;
```

we get the following result:



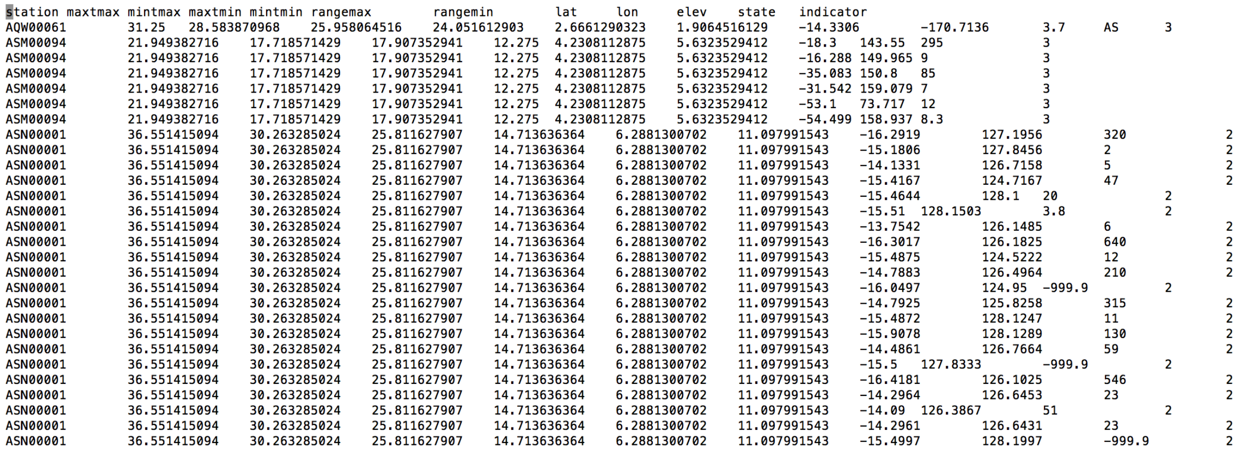
```
Obs    rangemax10    rangemin10

 1       6.68495       6.54762
```

so we can see that the boundary value for these two range is 6.68495 and 6.54762
then we can use this two value to determine the indicator varibales.
insert the following code in part(a) code:

```
data stations;
infile "ghcnd-stations.txt";
input station $ 1-11 lat 13-20 lon 22-30 elev 32-37 state $ 39-40;

proc sort data=stations out=stations2;
by station;

data data8;
merge data3(in=x) stations2(in=y);
by station;
if x=1 and y=1;

data data9;
set data8;
if rangemax ge 6.68495 and rangemin ge 6.54762 then
indicator=0;
else if rangemax ge 6.68495 and rangemin lt 6.54762 then
indicator=1;
else if rangemax lt 6.68495 and rangemin ge 6.54762 then
indicator=2;
else indicator=3;
if indicator=0 then delete;

proc export data=data9
dbms=tab
```

```
outfile='resultb.txt'
replace;
```

we get the following resultb.txt file, which will be imported to R to do further analysis, i post a few part of this txt file:



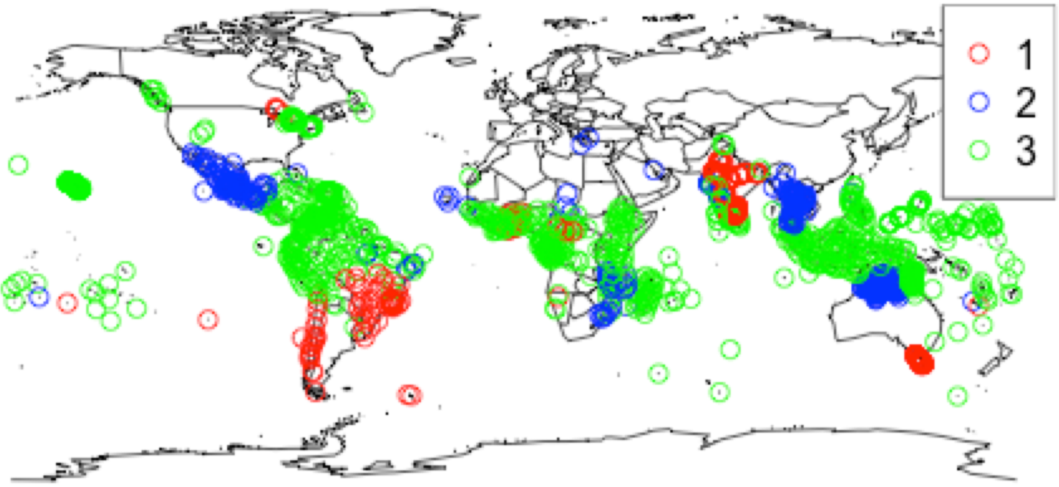
```
station maxtmax mintmax maxtmin mintmin rangemax        rangemin        lat     lon     elev    state   indicator
AQW00061        31.25   28.583870968    25.958064516    24.051612903    2.6661290323    1.9064516129    -14.3306        -170.7136       3.7     AS      3
ASM00094        21.949382716    17.718571429    17.907352941    12.275  4.2308112875    5.6323529412    -18.3   143.55  295             3
ASM00094        21.949382716    17.718571429    17.907352941    12.275  4.2308112875    5.6323529412    -16.288 149.965 9               3
ASM00094        21.949382716    17.718571429    17.907352941    12.275  4.2308112875    5.6323529412    -35.083 150.8   85              3
ASM00094        21.949382716    17.718571429    17.907352941    12.275  4.2308112875    5.6323529412    -31.542 159.079 7               3
ASM00094        21.949382716    17.718571429    17.907352941    12.275  4.2308112875    5.6323529412    -53.1   73.717  12              3
ASM00094        21.949382716    17.718571429    17.907352941    12.275  4.2308112875    5.6323529412    -54.499 158.937 8.3             3
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -16.2919        127.1956        320             2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.1806        127.8456        2               2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -14.1331        126.7158        5               2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.4167        124.7167        47              2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.4644        128.1   20              2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.51  128.1503        3.8             2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -13.7542        126.1485        6               2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -16.3017        126.1825        640             2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.4875        124.5222        12              2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -14.7883        126.4964        210             2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -16.0497        124.95  -999.9          2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -14.7925        125.8258        315             2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.4872        128.1247        11              2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.9078        128.1289        130             2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -14.4861        126.7664        59              2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.5   127.8333        -999.9          2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -16.4181        126.1025        546             2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -14.2964        126.6453        23              2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -14.09  126.3867        51              2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -14.2961        126.6431        23              2
ASN00001        36.551415094    30.263285024    25.811627907    14.713636364    6.2881300702    11.097991543    -15.4997        128.1997        -999.9          2
```

then we import this txt file into R to do the next analysis:

R code:

```
data=read.csv('resultb.txt',header=T,sep="\t")
library(mapproj)
library(maptools)
map()
coord=mapproject(data$lon,data$lat)
data$indicator=as.factor(data$indicator)
points(coord,col=c("red","blue","green")[data$indicator])
legend(x="topright", legend = levels(data$indicator),
col=c("red","blue","green"), pch=1)
```

and we get the following plot:



comment:

we know that:
indicator 3:both rangemax and rangemin are in bottom 10%
indicator 2:rangemax in bottom 10% while rangemin not
indicator 1:rangemin in bottom 10% while rangemin not

so from the plot, we can see that the place around equatorial, both rangemax and rangemin don't change much.

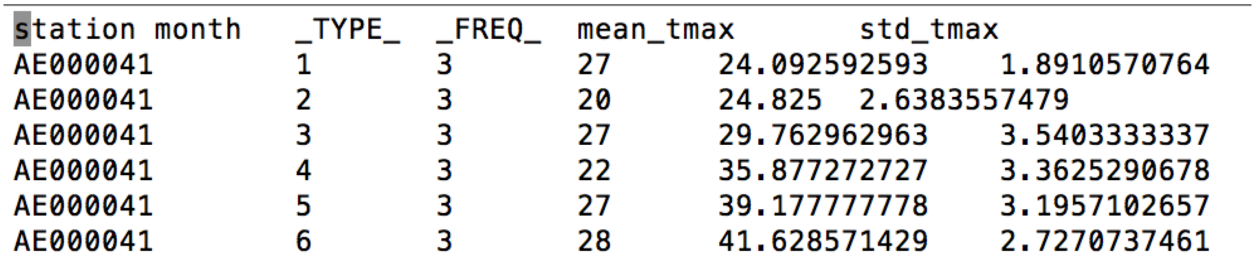
Problem2:

sas step:

code:

```
filename ghcnd_gz pipe "gzip -dc 2014.csv.gz" lrecl=80;

data ghcnd(rename=(obsval=tmax));
infile ghcnd_gz delimiter=",";
input station $ date : yymmdd8. obstype $ obsval;
format date mmddyy10.;
month = month(date);
if obstype = "TMAX";
obsval = obsval/10;

proc summary data=ghcnd(drop=obstype) nway;
class station month;
output out=ghcnd2
       mean(tmax)= mean_tmax
       std(tmax)=std_tmax;

proc export data=ghcnd2
dbms=tab
outfile='resultc.txt'
replace;
```

then we get the resultc.txt for later analysis in R, i post a few lines of resultc.txt:



```
station month   _TYPE_  _FREQ_  mean_tmax       std_tmax
AE000041        1       3       27      24.092592593    1.8910570764
AE000041        2       3       20      24.825  2.6383557479
AE000041        3       3       27      29.762962963    3.5403333337
AE000041        4       3       22      35.877272727    3.3625290678
AE000041        5       3       27      39.177777778    3.1957102657
AE000041        6       3       28      41.628571429    2.7270737461
```

then i import this file to R to do the analysis:

first we calculate the correlation between these two variable:

R code:

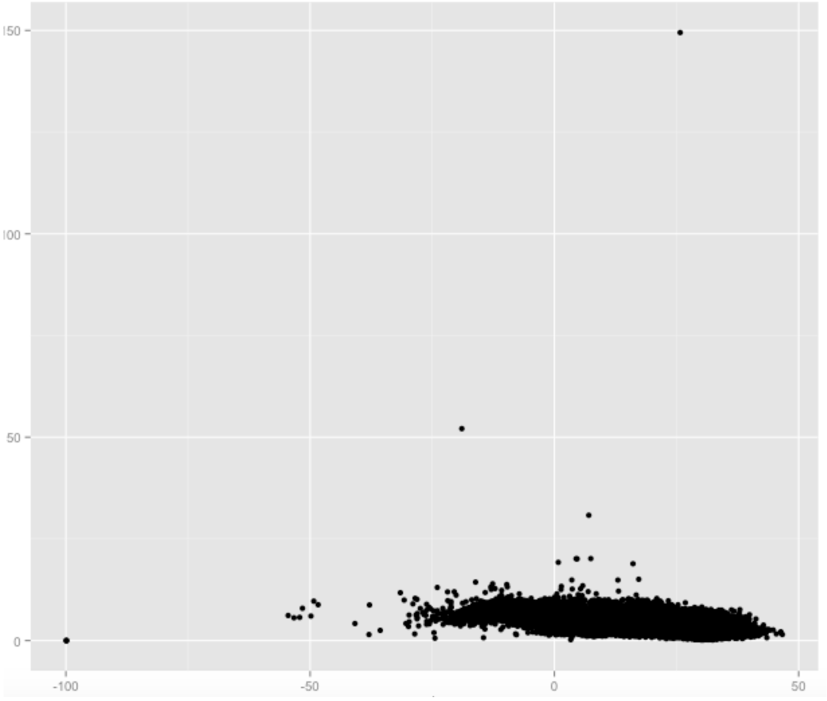
```
data1=na.omit(data1)
cor(data1$mean_tmax,data1$std_tmax)
```

and get the result:-0.5063753
so there is a negative relationship between these two variables, to make this more clear and plot these two variables.

R code:

```
data1=read.csv('resultc.txt',header=T,sep="\t")
library(ggplot2)
qplot(mean_tmax,std_tmax,data=data1)
```

we get the following plot:



and we can see that as n get larger, std tends to get smaller.

Problem 3:

(a)

code:

```
libname mydata '~/sas/';

filename ghcnd_gz pipe "gzip -dc 2013.csv.gz" lrecl=80;

data ghcnd(rename=(obsval=tmax2013));
infile ghcnd_gz delimiter=",";
input station $ date : yymmdd8. obstype $ obsval;
format date mmddyy10.;
month = month(date);
if obstype = "TMAX";
obsval = obsval/10;
if month = 1;
day=day(date);
keep station day obsval;

filename ghcn_gz pipe "gzip -dc 2014.csv.gz" lrecl=80;

data ghcnd2(rename=(obsval=tmax2014));
infile ghcn_gz delimiter=",";
input station $ date : yymmdd8. obstype $ obsval;
format date mmddyy10.;
month = month(date);
if obstype = "TMAX";
obsval = obsval/10;
if month = 1;
day=day(date);
keep station day obsval;

proc sort data=ghcnd out=ghcnd3;
by station;

proc sort data=ghcnd2 out=ghcnd4;
by station;

proc summary data=ghcnd3 nway;
class station day;
output out=ghcnd5(drop=_TYPE_ _FREQ_)
       mean(tmax2013)=mean_tmax2013;

proc summary data=ghcnd4 nway;
class station day;
output out=ghcnd6(drop=_TYPE_ _FREQ_)
       mean(tmax2014)=mean_tmax2014;
```

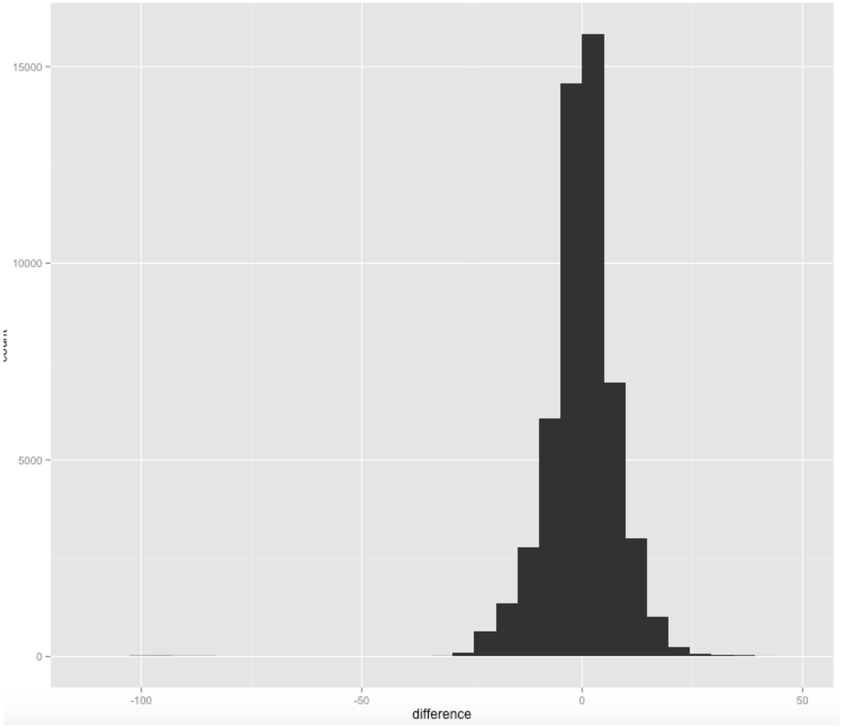
```
data ghcnd7;
merge ghcnd5(in=x) ghcnd6(in=y);
by station day;
if x=1 and y=1;

data ghcnd8;
set ghcnd7;
difference=mean_tmax2014−mean_tmax2013;

proc export data=ghcnd8
dbms=tab
outfile='resultd.txt'
replace;
```

then we get resultd.txt for the R analysis.

R code:

```
data2=read.csv('resultd.txt',header=T,sep="\t")
data2=na.omit(data2)
qplot(difference,data=data2,geom=“histogram")
```

and we get the following plot:



(b) **insert the following code in part(a):**

```
proc univariate data=ghcnd8;
var difference;
output out=ghcn9 pctlpts=10,90  pctlpre=difference;
```

we get the following result:

```
Quantile          Estimate

100% Max         43.800000
99%              18.183333
95%              12.033333
90%               8.904706
75% Q3            4.000000
50% Median        0.148214
25% Q1           -3.786154
10%              -9.366667
5%              -13.606250
1%              -21.014286
0% Min         -103.000000
```

so the boundary value of the top and bottom 10% of the distribution is 8.904706 and -9.366667

then we insert the following code in part(a):

```
data ghcnd9;
set ghcnd8;
if difference lt 8.904706 and  difference ge -9.366667 then delete;

data stations;
infile "ghcnd-stations.txt";
input station $ 1-11 lat 13-20 lon 22-30 elev 32-37 state $ 39-40;

proc sort data=stations out=stations2;
by station;

data ghcnd10;
merge ghcnd9(in=x) stations2(in=y);
by station;
if x=1 and y=1;

proc export data=ghcnd10
dbms=tab
```

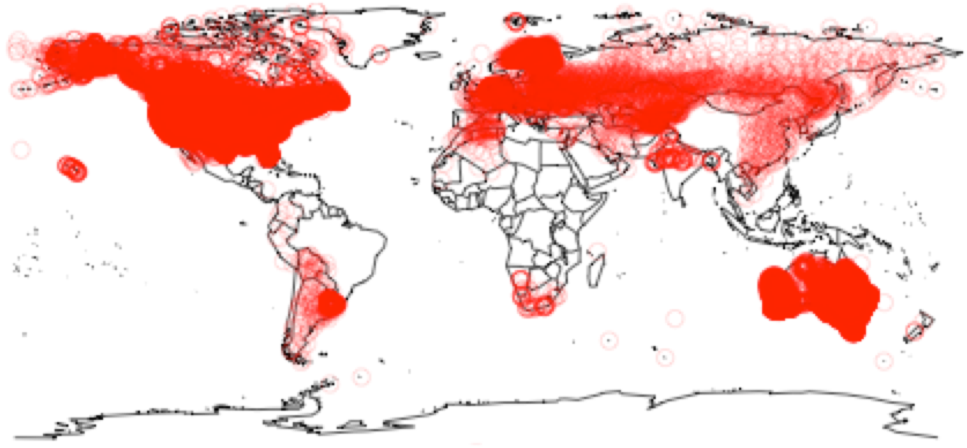
```
outfile='resulte.txt'
replace;
```

then we use R to map the data:

R code:

```
data3=read.csv('resulte.txt',header=T,sep="\t")
data3=na.omit(data3)
map()
coord=mapproject(data3$lon,data3$lat)
points(coord,col=rgb(1, 0, 0, 0.2))
```

we get the following plot:



see from the plot,we can see that change which is not extreme is not around equatorial.