```
/*import the dataset in climate in SAS  */

PROC IMPORT DATAFILE='/home/u45187342/my_courses/rafaeldeandrade0/ST662_data/climate.csv'
    out= st662.climate
    DBMS=CSV replace;
    GETNAMES=YES;
RUN;

/* identifying the missing value air_mean */

proc freq data=st662.climate;
table air_mean /nocum nopercent;
run;

/* identifying the missing values in air_mean */
proc sql;
create table st662.climate1 as
select site,year,count(*) as missing
from st662.climate where air_mean=. group by site,year;
quit();

/* identifying the freq of each site each year  */
proc sql;
create table st662.climate2 as
select site,year,count(*) as N
from st662.climate group by site,year;
quit();


/* sorting both the tables  */
proc sort data=st662.climate1;
by year;
run;
proc sort data=st662.climate2;
by year;
run;

/* merging the data set*/

data st662.climate12;
  merge st662.climate2 st662.climate1;
  by year site;
 run;

/* final dataset for question 1 reduced dataset*/
proc sql;
create table st662.climate11 as
select *,(missing/n)*100 as percent_missing from st662.climate12 where missing>0;
quit();

proc print data=st662.climate11;
run;
/* question 2 */

proc sort data=st662.climate11;
by site year;
run;
proc sort data=st662.climate;
by site year;
run;
data ST662.climate_2a;
```

```
merge ST662.climate11 ST662.climate;
by site year;
run;

proc sort data = ST662.climate_2a;
by site year;
run;

data  ST662.climate_2a;
set ST662.climate_2a;
by site year;
if percent_missing > 5  and air_min = . then air_mean = air_max/2;
if percent_missing > 5 and air_max=. then air_mean = air_min/2;
if percent_missing > 5 then air_mean = (air_min+air_max)/2;
run;

proc sort data = ST662.climate_2a;
by site year;
run;


proc stdize data= ST662.climate_2a out=ST662.climate_2a
            method= mean missing = mean reponly;
            by site year;
    var air_mean;
run;

/* question 3 */
proc sort data=ST662.climate11;
by site year;
run;

data st662.climatefinal2;
    merge  ST662.climate ST662.climate11;
    by site year;
run;

proc means data = ST662.climatefinal2 mean;
    by site year;
    var air_mean;
    where percent_missing >0;
output out = ST662.climateoldmean mean(air_mean) = old_mean;
run;

proc means data = ST662.climate_2a mean;
    by site year;
    var air_mean;
    where percent_missing >0;
    output out = ST662.climatenewmean mean(air_mean) = new_mean;
run;

data ST662.climatemeanfinal;
    merge ST662.climateoldmean ST662.climatenewmean;
    drop type freq;
run;

title"Reduced dataself for question 1 ";
proc print data=st662.climate11;
run;
title"Final dataset for question 3 ";
proc print data=ST662.climatemeanfinal;
run;
```

```
The old mean and new mean are same for each site and year,
Except for the missing percent is 100 there is no old mean(prior updating old mean).
```