```
/*A grassland biodiversity experiment was conducted at many sites across Europe and one in Canada. The
data from this experiment was published in the journal called Ecology. Information on the experiment
is available

1.Write a SAS programme to do the following data manipulation exercises.
1. (a) Download the biomass.csv dataset and read it into SAS.*/
PROC IMPORT OUT=st662.biomass
     DATAFILE='/home/u45187342/ST662/biomass.csv'
     DBMS=CSV replace;
     GETNAMES=YES;
RUN;


/*(b) Restrict the dataset to only sites 13, 14, 23, 25, 33 and 52, to only the first year of experimental
data, and to only treatment 1.*/

proc sql;
 select * from st662.biomass where site in(13,14,23,25,33,52) and yearn=1 and treat=1;
run;


/*(c) Create a new dataset that provides the annual yield for each plot at each site.*/

proc sql;
    create table st662.biomass1 as
 select year,site,plot,sum(harv_yield) as annual_yield
 from st662.biomass where site in(13,14,23,25,33,52) and yearn=1 and treat=1
 group by site,plot,year ;
quit;


proc print data=st662.biomass1;
run;


/*(d) Create a new dataset that provides the average annual yield for each site (i.e. averaged across
all plots).*/

proc sql;
    create table st662.biomass2 as
 select year,site,avg(Annual_yield) as Average_annual_yield
 from st662.biomass1 group by site,year ;
quit;


/*2. (a) Download the climate.csv dataset and read it into SAS.*/
PROC IMPORT OUT=st662.climate
     DATAFILE='/home/u45187342/my_courses/rafaeldeandrade0/ST662_data/climate.csv'
     DBMS=CSV replace;
     GETNAMES=YES;
RUN;

/*(b) Restrict the dataset to only sites 13, 14, 23, 25, 33 and 52.*/

proc sql;
 select * from st662.climate where site in(13,14,23,25,33,52);
run;

/*(c) Create a new dataset that provides the average 'air mean' for each site and each year.*/

proc sql;
    create table st662.climate1 as
 select year,site,avg(AIR_MEAN) as Average_annual_temperature
 from st662.climate
 where site in(13,14,23,25,33,52)
 group by site,year ;
quit;


/*3. (a) Merge the biomass dataset created in Qu 1d with the relevant year of the climate dataset
```

```sas
created in Qu 2c.*/

proc sort data=st662.biomass2;
by year;
run;
proc sort data=st662.climate1;
by year;
run;
data biomass_climate;
  merge st662.biomass2 st662.climate1;
  by year site;
 run;


data biomass_climate;
    set biomass_climate;
    where average_annual_yield ne .;
run;

/*(b) Create a scatter plot of average annual yield versus average annual temperature. Ensure the
quality of the scatterplot is suitable for including in a presentation or report (e.g. put a title
on it, check the font sizes of labels, perhaps label points within the graph etc).*/


proc sgplot data=biomass_climate;
  title'Average annual yield vs Average annual temperature';
  scatter x=average_annual_temperature y=average_annual_yield / datalabel = site;
  run;
```