```
/*************************** SAS Assingment-2 ***************************************/
/*Name:- Alok Kumar Singh                                                            */
/*ID:-   19250990                                                                    */
/************************************************************************************/


/*Q1 a) Download the biomass.csv dataset and read it into SAS. */
proc import out= ST662Lib.biomass
datafile= "/courses/d77u30vavpRs0h7u2Ms92/ST662_data/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.*/

data ST662Lib.biomass1;
set ST662Lib.biomass;
if site in (13, 14, 23, 25, 33, 52) and yearn = 1 and treat = 1 then output;
run;

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

proc freq data= ST662Lib.biomass1;
tables site plot / nocum nopercent;
run;

proc sort data = st662lib.biomass1;
    by site plot year;
run;

proc means data = st662lib.biomass1;
    by site plot year;
    var harv_yield;
    output out = st662lib.biomass2 sum(harv_yield) = annual_yield;
run;

/* d) Create a new dataset that provides the average annual yield for each site (i.e. averaged
across all plots).*/
proc means data = st662lib.biomass2;
    by site year;
    var annual_yield;
    output out = st662lib.biomass3 mean(annual_yield) = avg_annual_yield;
run;

/* Q2 a) Download the climate.csv dataset and read it into SAS.*/
proc import out= ST662Lib.climate
datafile= "/courses/d77u30vavpRs0h7u2Ms92/ST662_data/climate.csv"
dbms=csv replace;
getnames= yes;
run;

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

data st662lib.climate1;
    set st662lib.climate;
    if site in (13,14,23,25,33,52) then output;
run;
```

```
/* c) Create a new dataset that provides the average `air mean' for each site and each year.*/
proc sort data = st662lib.climate1;
    by site year;
run;
proc means data = st662lib.climate1;
    by site year;
    var air_mean;
    output out = st662lib.climate2 mean(air_mean) = avg_air_mean;
run;


/* Q3 a) Merge the biomass dataset created in Qu 1d with the relevant year of the climate
dataset created in Qu 2c.*/
proc sort data = st662lib.biomass3;
    by site year;
run;
proc sort data = st662lib.climate2;
    by site year;
run;

data st662lib.bmclfinal;
    merge st662lib.biomass3 st662lib.climate2;
    by site year;
    drop _freq_ _type_;
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).*/
data st662lib.bmclfinal;
    set st662lib.bmclfinal;
    if avg_annual_yield <> . then output;
run;
title1 "Data for Average Annual Yield vs Annula Temperature";
title2 " after removing the rows, where Average Annual Yield was missing";
proc print data = st662lib.bmclfinal;
run;

title "Average Annual Yield vs Average Annual Temperature";
proc sgplot data = st662lib.bmclfinal;
    scatter x = avg_air_mean y= avg_annual_yield / datalabel = site;
    label avg_air_mean = 'Average Annual Temperature' avg_annual_yield = 'Average Annual Yield';
run;
```