libname rawdata "C:\Users\huong.trinh\Desktop\household\_power\_consumption\datatosas";

**\*Import data;**

**proc** **import**

datafile="C:\Users\huong.trinh\Desktop\household\_power\_consumption\household\_power\_consumption\_1.csv"

out=household1 dbms=dlm

replace;

delimiter=",";

getnames=yes;

guessingrows=**20000**;

**run**;

**proc** **import**

datafile="C:\Users\huong.trinh\Desktop\household\_power\_consumption\household\_power\_consumption\_2.csv"

out=household2 dbms=dlm

replace;

delimiter=",";

getnames=yes;

guessingrows=**20000**;

**run**;

**data** group;

set household1 household2;

/\*Transform the values into numeric\*/

Global\_active\_power\_n=input(Global\_active\_power,??best.);

Global\_reactive\_power\_n=input(Global\_reactive\_power,??best.);

Voltage\_n=input(Voltage,??best.);

Global\_intensity\_n=input(Global\_intensity,??best.);

Sub\_metering\_1\_n=input(Sub\_metering\_1,??best.);

Sub\_metering\_2\_n=input(Sub\_metering\_2,??best.);

Sub\_metering\_3\_n=input(Sub\_metering\_3,??best.);

drop Global\_active\_power Global\_reactive\_power Voltage Global\_intensity Sub\_metering\_1 Sub\_metering\_2 Sub\_metering\_3;

rename Global\_active\_power\_n=Global\_active\_power

Global\_reactive\_power\_n=Global\_reactive\_power Voltage\_n=Voltage Global\_intensity\_n=Global\_intensity

Sub\_metering\_1\_n=Sub\_metering\_1 sub\_metering\_2\_n=Sub\_metering\_2 sub\_metering\_3\_n=Sub\_metering\_3;

**run**;

**data** rawdata.rawpower;

set group;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**data** rawpower;

set rawdata.rawpower;

/\*Try to make the first plot as time series based on all data \*/

datetime=dhms(Date, **0**, **0**, Time);

format datetime datetime23.;

monthcat=month(date);

daycat=day(date);

yearcat=year(date);

hourcat=hour(time);

other=Global\_active\_power\***1000**/**60** - Sub\_metering\_1-Sub\_metering\_2-Sub\_metering\_3;

**proc** **sort**;

by date hourcat;

**run**;

**proc** **sql**;

create table rawpower1 as

select date, yearcat, monthcat, daycat, hourcat, sum(Global\_active\_power) \***1000**/**60** as sum\_active, sum(global\_reactive\_power) \***100**/**60** as sum\_reactive,

sum(Sub\_metering\_1) as sum\_sub1, sum(sub\_metering\_2) as sum\_sub2, sum(sub\_metering\_3) as sum\_sub3,

sum(other) as sum\_other, avg(voltage)as avg\_vol, avg(global\_intensity) as avg\_int

from rawpower

group by date, yearcat, monthcat, daycat, hourcat

order by yearcat, monthcat, hourcat, daycat, date;

**quit**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*Now get the average by hourcat for every day in each month;

**proc** **sql**;

create table rawpower2 as

select yearcat, monthcat, hourcat, avg(sum\_active) as avg\_active1, avg(sum\_reactive) as avg\_reactive1, avg(sum\_sub1) as avg\_sub11,

avg(sum\_sub2) as avg\_sub21, avg(sum\_sub3) as avg\_sub31, avg(sum\_other) as avg\_other1, avg(avg\_vol) as avg\_vol1,

avg(avg\_int) as avg\_int1

from rawpower1

group by yearcat, monthcat, hourcat

order by monthcat, hourcat, yearcat;

**quit**;

\*Now get the average by year;

**proc** **sql**;

create table rawpower3 as

select monthcat, hourcat, avg(avg\_active1) as avg\_active2, avg(avg\_reactive1) as avg\_reactive2, avg(avg\_sub11) as avg\_sub12,

avg(avg\_sub21) as avg\_sub22, avg(avg\_sub31) as avg\_sub32, avg(avg\_other1) as avg\_other2, avg(avg\_vol1) as avg\_vol2,

avg(avg\_int1) as avg\_int2

from rawpower2

group by monthcat, hourcat

order by monthcat, hourcat ;

**quit**;

**data** rawpower3;

set rawpower3;

ordering=\_N\_;

label avg\_sub12="Monthly average of the averaged usage of the kitchen (containing mainly a dishwasher, an oven and a microwave) per hour block (0-23)"

avg\_sub22="Monthly average of the averaged usage of the laundry room (containing a washing-machine, a tumble-dryer, a refrigator and a light) per hour block (0-23)"

avg\_sub32="Monthly average of the averaged usage of the electric water-heater and the air-conditioner per hour block (0-23)"

avg\_other2="Monthly average of the averaged usage of other appliance at home per hour block (0-23)";

**proc** **sort**;

by monthcat hourcat;

**run**;

/\*Dummy data for hourcat to represent every hour block in each montht\*/

**data** hourcat;

set rawpower3;

keep hourcat;

**proc** **sort** nodupkey;

by hourcat;

**run**;

**data** monthcat;

set rawpower3;

keep monthcat;

**proc** **sort** nodupkey;

by monthcat;

**run**;

**proc** **sql**;

create table dummy as

select hourcat, monthcat

from monthcat, hourcat

order by monthcat, hourcat;

**quit**;

**data** rawpower4;

merge dummy rawpower3;

by monthcat hourcat;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

ods path show;

ods path(prepend) work.templat(update);

**proc** **template**;

define statgraph block;

begingraph/designwidth=**1200**px designheight=**800**px;

entrytitle 'Monthly Cycle of the 4-Year Based Monthly Average Usage of Active Power in Each Hour in Different Sub-metering Categories';

layout overlay/ xaxisopts=(type=linear offsetmin=**0.01** offsetmax=**0.01**

label="From left to right, each color block represents a month from January to December. In each color block, from left to right is hour category from 0 to 23"

display=all /\*(line label tickvalues)\*/ labelattrs = (size = **11**) )

yaxisopts=(type=linear offsetmin=**0.01** offsetmax=**0.01**

label ="Day Averaged Usage of Active Power in Each Month in Each Year" display = all labelattrs = (size = **11**));

blockplot x=ordering block=monthcat/valuevalign=bottom datatransparency=**0.75** display=(fill values);

scatterplot x=ordering y=avg\_sub12/ datalabel=hourcat markerattrs=(symbol=circle color=red size=**3**px);

seriesplot x=ordering y=avg\_sub12/name="sub1" lineattrs=(pattern=solid color=red);

scatterplot x=ordering y=avg\_sub22/ markerattrs=(symbol=plus color=blue size=**3**px);

seriesplot x=ordering y=avg\_sub22/name="sub2" lineattrs=(pattern=shortdash color=blue);

scatterplot x=ordering y=avg\_sub32/ datalabel=hourcat markerattrs=(symbol=diamond color=green size=**3**px);

seriesplot x=ordering y=avg\_sub32/name="sub3" lineattrs=(pattern=dot color=green);

scatterplot x=ordering y=avg\_other2/ datalabel=hourcat markerattrs=(symbol=Square color=purple size=**3**px);

seriesplot x=ordering y=avg\_other2/name="subother" lineattrs=(pattern=MediumDash color=purple);

discretelegend "sub1" "sub2" "sub3" "subother"/ location=outside halign=left valign=bottom border=off across=**1**;

endlayout;

endgraph;

end;

**run**;

**proc** **sgrender** data=rawpower4 template=block;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*Ok, make another bar plot for total sum \*/

**proc** **sql**;

create table rawpower5 as

select yearcat, monthcat, sum(sum\_active)/**1000** as sum\_active1, sum(sum\_reactive)/**1000** as sum\_reactive1,

sum(sum\_sub1)/**1000** as sum\_sub11, sum(sum\_sub2)/**1000** as sum\_sub21, sum(sum\_sub3)/**1000** as sum\_sub31,

sum(sum\_other)/**1000** as sum\_other1, avg( avg\_vol)as avg\_vol1, avg(avg\_int) as avg\_int1

from rawpower1

group by yearcat, monthcat

order by yearcat, monthcat;

**quit**;

**proc** **transpose** data=rawpower5 out=rawpower5\_tp;

by yearcat monthcat;

var sum\_sub11 sum\_sub21 sum\_sub31 sum\_other1 sum\_reactive1;

**run**;

**data** rawpower5\_tp;

set rawpower5\_tp;

length cat $**200**;

select(\_name\_);

when ("sum\_sub11")

cat="Total active energy consumed by the kitchen (containing mainly a dishwasher, an oven and a microwave";

when ("sum\_sub21")

cat="Total active energy consumed by the laundry room (containing a washing-machine, a tumble-dryer, a refrigator and a light";

when ("sum\_sub31")

cat="Total active energy consumed by the electric water-heater and an air-conditioner";

when ("sum\_other1")

cat="Total active energy consumed by other appliance";

otherwise

cat="Total reactive energy";

end;

label cat="Category" col1="Total energy consumed in a month (Watt-hour)";

**run**;

ods graphics on/width=**16**in height=**8**in;

title "Total Energy Consumed by Month, Year and Type";

**proc** **sgpanel** data=rawpower5\_tp;

panelby yearcat/layout=columnlattice onepanel novarname noborder colheaderpos=bottom;

vbar monthcat/ barwidth=**1** response=col1 group=cat groupdisplay=cluster;

colaxis display=(nolabel);

rowaxis grid;

**run**;