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

PURPOSE: Final Project

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

libname mylib '/home/u63669078/final project';

\* assigns library to tell SAS where data sets are, BASED ON WHERE YOU SAVED YOUR DATA SET(s);

\* Print all variables and all observations in 8 data sets;

PROC PRINT DATA=MYLIB.CD0531;

RUN;

PROC PRINT DATA=MYLIB.CD0601;

RUN;

PROC PRINT DATA=MYLIB.CD0602;

RUN;

PROC PRINT DATA=MYLIB.CD0603;

RUN;

PROC PRINT DATA=MYLIB.CD0604;

RUN;

PROC PRINT DATA=MYLIB.CD0605;

RUN;

PROC PRINT DATA=MYLIB.CD0606;

RUN;

PROC PRINT DATA=MYLIB.CD0607;

RUN;

\* Get the subset of the data set on June 1 only for State=Delaware;

data cd0601\_1;

set mylib.cd0601;

where Province\_State="Delaware";

run;

\* Get the subset of the data set on June 7 only for State=Delaware;

data cd0607\_1;

set mylib.cd0607;

where Province\_State="Delaware";

run;

\*Firstly, we need to get the subset data for Delaware from May 31 to June 7, 2020;

data cd0531\_1;

set mylib.cd0531;

where Province\_State="Delaware";

run;

data cd0602\_1;

set mylib.cd0602;

where Province\_State="Delaware";

run;

data cd0603\_1;

set mylib.cd0603;

where Province\_State="Delaware";

run;

data cd0604\_1;

set mylib.cd0604;

where Province\_State="Delaware";

run;

data cd0605\_1;

set mylib.cd0605;

where Province\_State="Delaware";

run;

data cd0606\_1;

set mylib.cd0606;

where Province\_State="Delaware";

run;

\* Before merging all data sets, we need to rename each variables in each data set to prevent overlap of data;

data cd0531\_2;

set cd0531\_1;

Confirmed1=Confirmed;

Deaths1=Deaths;

Tested1=People\_Tested;

Hospitalized1=People\_Hospitalized;

run;

data cd0601\_2;

set cd0601\_1;

Confirmed2=Confirmed;

Deaths2=Deaths;

Tested2=People\_Tested;

Hospitalized2=People\_Hospitalized;

run;

data cd0602\_2;

set cd0602\_1;

Confirmed3=Confirmed;

Deaths3=Deaths;

Tested3=People\_Tested;

Hospitalized3=People\_Hospitalized;

run;

data cd0603\_2;

set cd0603\_1;

Confirmed4=Confirmed;

Deaths4=Deaths;

Tested4=People\_Tested;

Hospitalized4=People\_Hospitalized;

run;

data cd0604\_2;

set cd0604\_1;

Confirmed5=Confirmed;

Deaths5=Deaths;

Tested5=People\_Tested;

Hospitalized5=People\_Hospitalized;

run;

data cd0605\_2;

set cd0605\_1;

Confirmed6=Confirmed;

Deaths6=Deaths;

Tested6=People\_Tested;

Hospitalized6=People\_Hospitalized;

run;

data cd0606\_2;

set cd0606\_1;

Confirmed7=Confirmed;

Deaths7=Deaths;

Tested7=People\_Tested;

Hospitalized7=People\_Hospitalized;

run;

data cd0607\_2;

set cd0607\_1;

Confirmed8=Confirmed;

Deaths8=Deaths;

Tested8=People\_Tested;

Hospitalized8=People\_Hospitalized;

run;

\*Sort the data by the unique identifier;

proc sort data = cd0531\_2; by Province\_State; run;

proc sort data = cd0601\_2; by Province\_State; run;

proc sort data = cd0602\_2; by Province\_State; run;

proc sort data = cd0603\_2; by Province\_State; run;

proc sort data = cd0604\_2; by Province\_State; run;

proc sort data = cd0605\_2; by Province\_State; run;

proc sort data = cd0606\_2; by Province\_State; run;

proc sort data = cd0607\_2; by Province\_State; run;

\* Merge 8 datasets by Province\_state;

data merged1;

merge cd0531\_2 cd0601\_2 cd0602\_2 cd0603\_2 cd0604\_2 cd0605\_2 cd0606\_2 cd0607\_2;

by Province\_State;

run;

\* To view the output merged table;

proc print data=merged1;

run;

\* Calculate the new cases of Deaths in each day from June 1 to June 7;

data merged2;

set merged1;

\* new cases on June 1;

newcases1=Deaths2-Deaths1;

\* new cases on June 2;

newcases2=Deaths3-Deaths2;

\* new cases on June 3;

newcases3=Deaths4-Deaths3;

\* new cases on June 4;

newcases4=Deaths5-Deaths4;

\* new cases on June 5;

newcases5=Deaths6-Deaths5;

\* new cases on June 6;

newcases6=Deaths7-Deaths6;

\* new cases on June 7;

newcases7=Deaths8-Deaths7;

run;

\* To view the new merge table with new cases of Deaths;

proc print data=merged2;

run;

\* Keep newcases variable;

data merged3;

set merged2;

keep newcases1 newcases2 newcases3 newcases4 newcases5 newcases6 newcases7 Province\_State;

run;

\* Convert the table into two variables Date and Deaths;

data merged4;

set merged3;

array newcases newcases1-newcases7;

do Date = 1 to 7;

Deaths = newcases[Date];

output;

end;

drop newcases1-newcases7;

run;

\* Get the plot to visualise;

proc sgplot data=merged4;

scatter x=Date y=Deaths;

xaxis label="Date";

yaxis label="Number of New Deaths";

Title "Daily COVID-19 Deaths";

run;

\* To see which region is Delaware in ;

data region1;

set mylib.region;

where Province\_State="Delaware";

run;

\* Find all states in South;

data region2;

set mylib.region;

where Region="South";

run;

\*Then, sort and merge all datasets contains all states;

data cd0531\_5;

set mylib.cd0531;

Confirmed1=Confirmed;

Deaths1=Deaths;

Tested1=People\_Tested;

Hospitalized1=People\_Hospitalized;

run;

data cd0601\_5;

set mylib.cd0601;

Confirmed2=Confirmed;

Deaths2=Deaths;

Tested2=People\_Tested;

Hospitalized2=People\_Hospitalized;

run;

data cd0602\_5;

set mylib.cd0602;

Confirmed3=Confirmed;

Deaths3=Deaths;

Tested3=People\_Tested;

Hospitalized3=People\_Hospitalized;

run;

data cd0603\_5;

set mylib.cd0603;

Confirmed4=Confirmed;

Deaths4=Deaths;

Tested4=People\_Tested;

Hospitalized4=People\_Hospitalized;

run;

data cd0604\_5;

set mylib.cd0604;

Confirmed5=Confirmed;

Deaths5=Deaths;

Tested5=People\_Tested;

Hospitalized5=People\_Hospitalized;

run;

data cd0605\_5;

set mylib.cd0605;

Confirmed6=Confirmed;

Deaths6=Deaths;

Tested6=People\_Tested;

Hospitalized6=People\_Hospitalized;

run;

data cd0606\_5;

set mylib.cd0606;

Confirmed7=Confirmed;

Deaths7=Deaths;

Tested7=People\_Tested;

Hospitalized7=People\_Hospitalized;

run;

data cd0607\_5;

set mylib.cd0607;

Confirmed8=Confirmed;

Deaths8=Deaths;

Tested8=People\_Tested;

Hospitalized8=People\_Hospitalized;

run;

\*Sort the data by the unique identifier;

proc sort data = cd0531\_5; by Province\_State; run;

proc sort data = cd0601\_5; by Province\_State; run;

proc sort data = cd0602\_5; by Province\_State; run;

proc sort data = cd0603\_5; by Province\_State; run;

proc sort data = cd0604\_5; by Province\_State; run;

proc sort data = cd0605\_5; by Province\_State; run;

proc sort data = cd0606\_5; by Province\_State; run;

proc sort data = cd0607\_5; by Province\_State; run;

\* Merge 8 datasets by Province\_state;

data merged6;

merge cd0531\_5 cd0601\_5 cd0602\_5 cd0603\_5 cd0604\_5 cd0605\_5 cd0606\_5 cd0607\_5;

by Province\_State;

run;

\* Sort the Region data;

proc sort data = region2; by Province\_State; run;

\* Merge them together;

data merged10;

merge merged6 region2;

by Province\_State;

run;

\* Find States in South;

data merged11;

set merged10;

where Region="South";

run;

\* Check the output;

proc print data=merged11;

run;

\* Keep necessary outcomes on June 1;

data merged12;

set merged11;

keep Province\_state Deaths2 Region;

run;

\* Get mean of Deaths on June 1 in South;

proc means data=merged12;

var Deaths2;

run;

\* Keep necessary outcomes on June 7;

data merged13;

set merged11;

keep Province\_state Deaths8 Region;

run;

\* Get mean of Deaths on June 7 in South;

proc means data=merged13;

var Deaths8;

run;

\* Get the name with the highest Deaths at the end day of the study;

data merged14;

set merged13;

where Deaths8=2936;

run;

\* Get the name with the lowest Deaths at the end day of the study;

data merged15;

set merged13;

where Deaths8=84;

run;