```
/*-------------------------------------------------------------------------------------------------------------------------------------------------------
Contents
    1) Satterplots: Positively Associated, Negatively Associated, Non-linearly Associated (With & Wihout Outlier)
    2) Correlations: Positive, Negative (With & Wothout Outlier)
    3) Regression Line: Fitted Regression line with & without outlier
    4) Two Way Table: Coross Tabulation, Marginal Proportions, Conditional Distribution
    5) Lab Assignmnet: Discussion
-------------------------------------------------------------------------------------------------------------------------------------------------------*/
/*=======================*
|Datasets: Overview |
*=======================*/

/* Sheet: pos_linear */

FILENAME REFFILE '/home/u64006622/Datasets.xlsx';
proc import datafile=reffile
    dbms=xlsx
    out=pos_linear
    replace;
    sheet="pos_linear";
    getnames=yes;
run;

/* Sheet : pos_linear_outlier */
proc import datafile=reffile
    dbms=xlsx
    out=pos_linear_outlier
    replace;
    sheet="pos_linear_outlier";
    getnames=yes;
run;

/* Sheet : neg_linear */
proc import datafile=reffile
    dbms=xlsx
    out=neg_linear
    replace;
    sheet="neg_linear";
    getnames=yes;
run;

/* Sheet : neg_linear_outlier */
proc import datafile=reffile
    dbms=xlsx
    out=neg_linear_outlier
    replace;
    sheet="neg_linear_outlier";
    getnames=yes;
run;

/* Sheet : nonlinear */
proc import datafile="reffile"
    dbms=xlsx out=nonlinear replace;
    sheet="nonlinear";
    getnames=yes;
run;

/* Sheet : nonlinear_outlier */
proc import datafile="reffile"
    dbms=xlsx out=nonlinear_outlier replace;
```

```
    sheet="nonlinear_outlier";
    getnames=yes;
run;
/*-------------------------------------------------------------------------------------------------------------------------------------------------------------*/

/*=======================*
|1) Scatterplots         |
*=======================*/

/* Positive association (without and with outliers) */
/* Without Outliers */
title "Scatter: Positive Linear Association (no outlier)";
proc sgplot data=pos_linear;
  scatter x=x y=y;
run;

/*With Outliers */
title "Scatter: Positive Linear Association (with outlier)";
proc sgplot data=pos_linear_outlier;
  scatter x=x y=y;
run;


/* Negative association without outliers */
title "Scatter: Negative Linear Association (no outlier)";
proc sgplot data=neg_linear;
  scatter x=x y=y;
run;

/* Negative association with outliers */
title "Scatter: Negative Linear Association (with outlier)";
proc sgplot data=neg_linear_outlier;
  scatter x=x y=y;
run;

/* Non-linear association without outliers */
title "Scatter: Non-linear (Quadratic) (no outlier)";
proc sgplot data=nonlinear;
  loess x=x y=y / nomarkers;  /* smooth curve to highlight non-linearity */
  scatter x=x y=y;
run;

/* Non-linear association with outliers */
title "Scatter: Non-linear (Quadratic) (with outlier)";
proc sgplot data=nonlinear_outlier;
  loess x=x y=y / nomarkers;
  scatter x=x y=y;
run;


/*-------------------------------------------------------------------------------------------------------------------------------------------------------------*/

/*=======================*
|2) Correlations         |
*=======================*/

/* Positive correlation (with and without outlier) */
title "Correlation: Positive (no outlier)";
proc corr data=pos_linear plots=scatter;
  var x y;
run;
```

```
title "Correlation: Positive (with outlier)";
proc corr data=pos_linear_outlier plots=scatter;
  var x y;
run;

/* Negative correlation (with and without outlier) */
title "Correlation: Negative (no outlier)";
proc corr data=neg_linear plots=scatter;
  var x y;
run;

title "Correlation: Negative (with outlier)";
proc corr data=neg_linear_outlier plots=scatter;
  var x y;
run;

/*---------------------------------------------------------------------------------------------------------------------------------------------------------------*/

/*===========================*
|3) Least-Squares Regression |
*===========================*/

/* Fitted regression line on a scatterplot */
title "Least-Squares Regression Line (Positive Linear Data)";
proc sgplot data=pos_linear;
  reg x=x y=y;     /* draws the regression line */
  scatter x=x y=y;
run;


/* With outlier */
title "Least-Squares Regression Line (Positive Linear Data with outlier)";
proc sgplot data=pos_linear_outlier;
  reg x=x y=y;     /* draws the regression line */
  scatter x=x y=y;
run;

/*---------------------------------------------------------------------------------------------------------------------------------------------------------------*/
/*==================*
|4) Two-way Table |
*==================*/
/*Suppose we have students classified by Gender (Male/Female) and Grade (Pass/Fail):*/
data students;
    input Gender $ Grade $ Count;
    datalines;
Male Pass 15
Male Fail 5
Female Pass 18
Female Fail 2
;
run;

/* 4A) Two-way table (cross-tabulation) */
proc freq data=students;
    tables Gender*Grade / norow nocol nopercent;
    weight Count;
run;

/* 4B) Distribution plot of the two-way table */
proc sgplot data=students;
```

```
    vbar Gender / response=Count group=Grade groupdisplay=cluster;
    xaxis label="Gender";
    yaxis label="Number of Students";
    title "Two-Way Table: Gender vs Grade";
run;

/* 4C) Two-way table with proportions */
proc freq data=students;
    tables Gender*Grade / nocol norow;
    weight Count;
run;

/* 4D) Marginal proportions  */
proc freq data=students;
    tables Gender Grade / nocum;
    weight Count;
run;

/* 4E)Conditional distribution table (Grade | Gender) */
proc freq data=students;
    tables Gender*Grade / nocol nopercent;
    weight Count;
run;

/* 4F) Conditional distribution plot: stacked proportions */
proc sgplot data=students;
    vbar Gender / response=Count group=Grade groupdisplay=stack stat=percent;
    xaxis label="Gender";
    yaxis label="Percent within Gender";
    title "Conditional Distribution of Grade Given Gender";
run;
```