CURVES AND SPLINES
DATA INTERPOLATION

- **SGPLOT** provides various methods for fitting “smooth” trends to scatterplot data
  - **LOESS**—An extension of LOWESS (Locally Weighted Scatterplot Smoothing), uses locally weighted polynomial regression.
  - **PBSPLINE**—Penalized B-splines. Piece-wise polynomials joined at “knots”; both the degree of the polynomial(s) and the penalty can be used to control smoothness.
  - **SPLINE**—Piece-wise Bézier curves. For these, continuity of derivatives to a given order is prescribed to create a smoothness condition.
BASIC LOESS, PBSPLINE & SPLINE

- Same basic syntax is common to all three:

  Title 'High Value in Year 2000';
  proc sgplot data=sashelp.stocks;
    loess x=date y=high;
    where date between '01JAN2000'd and '31DEC2000'd;
  run;
  quit;

  proc sgplot data=sashelp.stocks;
    pbspline x=date y=high;
    where date between '01JAN2000'd and '31DEC2000'd;
  run;
  quit;

  proc sgplot data=sashelp.stocks;
    spline x=date y=high;
    where date between '01JAN2000'd and '31DEC2000'd;
  run;
  quit;
BASIC LOESS, PBSPLINE & SPLINE

- Same basic syntax is common to all three:

```sas
Title 'High Value in Year 2000';
proc sgplot data=sashelp.stocks;
   loess x=date y=high;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;

proc sgplot data=sashelp.stocks;
   pbspline x=date y=high;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;

proc sgplot data=sashelp.stocks;
   spline x=date y=high;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```

Each requires specification of an x and y variable.
BASIC LOESS, PBSPLINE & SPLINE

- Same basic syntax is common to all three:

```plaintext
Title 'High Value in Year 2000';
proc sgplot data=sashelp.stocks;
   loess x=date y=high;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;

proc sgplot data=sashelp.stocks;
   pbspline x=date y=high;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;

proc sgplot data=sashelp.stocks;
   spline x=date y=high;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```

Uh-oh.
RESULTS

High Value in Year 2000

LOESS

B-spline
RESULTS

Spline is available in maintenance release 3 and up.
It is reasonable to expect some options to be common to both, and perhaps the same or similar to some previous options. One of those is group:

```sas
proc sgplot data=sashelp.stocks;
   loess x=date y=high / group=stock;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;

proc sgplot data=sashelp.stocks;
   pbspline x=date y=high / group=stock;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```
RESULTS

High Value in Year 2000

LOESS

B-spline
Since the curve plots include data markers by default, the options for those are almost identical to those covered in the SCATTER statement. They can also be removed:

```sas
proc sgplot data=sashelp.stocks;
    loess x=date y=high / group=stock nomarkers;
    where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;

proc sgplot data=sashelp.stocks;
    pbspline x=date y=high / group=stock nomarkers;
    where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```
RESULTS

**High Value in Year 2000**

- LOESS
- B-spline
The curves themselves are viewed as lines by SAS, so \texttt{LINEATTRS=} is useful here:

```sas
proc sgplot data=sashelp.stocks;
   loess y=high / group=stock nomarkers lineattrs=(pattern=1);
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;

proc sgplot data=sashelp.stocks;
   pbspline y=high / group=stock nomarkers lineattrs=(pattern=1);
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```
RESULTS
As usual, for groups SAS alters both color and pattern. You can set either, but only set one.
• Confidence limits for the mean are possible in both, and for individual predictions in PBSPLINE; however, there is currently a flaw that does not permit either with a group variable in PBSPLINE:

``` SAS 
proc sgplot data=sashelp.stocks;
   loess x=date y=high / group=stock nomarkers clm;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;

proc sgplot data=sashelp.stocks;
   pbspline x=date y=high / nomarkers clm;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```
RESULTS

High Value in Year 2000

- B-spline
- LOESS
RESULTS

Since these are bands, transparency can be set with CLMTRANSPARENCY=. 
· Since the curve fitting methods differ between the two types, controlling the method of smoothness does as well. In LOESS:

```sas
proc sgplot data=sashelp.stocks;
   loess x=date y=high / smooth=0.4 degree=2 interpolation=cubic;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```
Since the curve fitting methods differ between the two types, controlling the method of smoothness does as well. In LOESS:

```
proc sgplot data=sashelp.stocks;
    loess x=date y=high / smooth=0.4 degree=2 interpolation=cubic;
    where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```

Smooth picks the proportion of the data used to estimate the function at each bin.
CONTROLLING THE SMOOTHING

- Since the curve fitting methods differ between the two types, controlling the method of smoothness does as well. In LOESS:

```sas
proc sgplot data=sashelp.stocks;
   loess x=date y=high / smooth=0.4 degree=2 interpolation=cubic;
   where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```

Degree of the local regression used.

Smooth must be between \((\text{degree}+1)/n\) and 1.
Since the curve fitting methods differ between the two types, controlling the method of smoothness does as well. In LOESS:

```sas
proc sgplot data=sashelp.stocks;
    loess x=date y=high / smooth=0.4 degree=2 interpolation=cubic;
    where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```

Joining of local regressions. Help says default is cubic, but it is not...
SOME RESULTS

Degree=1,
Smooth=0.1
Linear

Degree=2,
Smooth=0.5
Cubic
SOME RESULTS

Low smoothness with linear regression and interpolation. Can see bins are distinct values of the x-variable.
Some Results

Degree=1, Smooth=0.1
Linear

Degree=2, Smooth=0.5
Cubic

Local quadratic and cubic interpolation produce a much less jagged construction. Greater smoothing value produces smoother curve.
CONTROLLING THE SMOOTHING

- Since the curve fitting methods differ between the two types, controlling the method of smoothness does as well. In LOESS:

```sas
proc sgplot data=sashelp.stocks;
    pbspline x=date y=high / smooth=0.1 degree=2;
    where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```
Since the curve fitting methods differ between the two types, controlling the method of smoothness does as well. In LOESS:

```plaintext
proc sgplot data=sashelp.stocks;
    pbspline x=date y=high / smooth=0.1 degree=2;
    where date between '01JAN2000'd and '31DEC2000'd;
run;
quit;
```

Same rules as before for the minimum on smooth. Degree can be any whole number between 0 and 10.
SOME RESULTS

Degree=2, Smooth=0.1

Degree=3, Smooth=10,000
SOME RESULTS

Degree=2, Smooth=0.1

Default degree is 3, but 2 is sufficient to avoid jagged connections even if the smoothing value is low.
SOME RESULTS

Smoothing value is not capped at 1 here (functions differently). Very large values may be required to remove “wiggles”.

\[
\text{Degree}=2, \quad \text{Smooth}=0.1
\]

\[
\text{Degree}=3, \quad \text{Smooth}=10,000
\]
EXERCISE 1 (USING CARS DATA)
EXERCISE 2 (STOCKS)