Exploratory Statistical Data Analysis With R Software (ESDAR)

Swayam Prabha

Lecture 34

Association of Variables : Smooth Scatter Plots

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from

http://home.iitk.ac.in/~shalab/sp

Scatter Plots with Smooth Curve

Suppose two variables are related.

A scatter plot along with a fitted line will provide information on the trend or relationship between them.

`scatter.smooth` produces a scatter plot and adds a smooth curve to the scatter plot.

Scatter Plots with Smooth Curve

`scatter.smooth` is based on the concept of LOESS which is a locally weighted scatterplot smoothing method.

LOESS is used for local polynomial regression fitting.

Fit a polynomial surface determined by one or more numerical predictors, using local fitting.

Use `help("scatter.smooth")` to get more details.

Scatter Plots with Smooth Curve

```
scatter.smooth(x, y = NULL, span = 2/3, degree
= 1, family = c("symmetric","gaussian"), xlab =
NULL, ylab = NULL, ylim = range(y, pred$y,
na.rm = TRUE),...)
```

`x, y`	`x` and `y` arguments provide the x and y coordinates for the plot.
`span`	smoothness parameter for LOESS.
`degree`	degree of local polynomial used.
`family`	if "gaussian" fitting is by least-squares, and if family = "symmetric" a re-descending M estimator is used.
`xlab`	label for x axis.
`ylab`	label for y axis.
`ylim`	the y limits of the plot.

Scatter Plots with Smooth Curve

Example

Data on marks obtained by 20 students out of 500 marks and the number of hours they studied per week are recorded as follows:

We know from experience that marks obtained by students increase as the number of hours increase.

Marks	337	316	327	340	374	330	352	353	370	380
Number of hours per week	23	25	26	27	30	26	29	32	33	34

Marks	384	398	413	428	430	438	439	479	460	450
Number of hours per week	35	38	39	42	43	44	45	46	44	41

Scatter Plots with Smooth Curve

Example

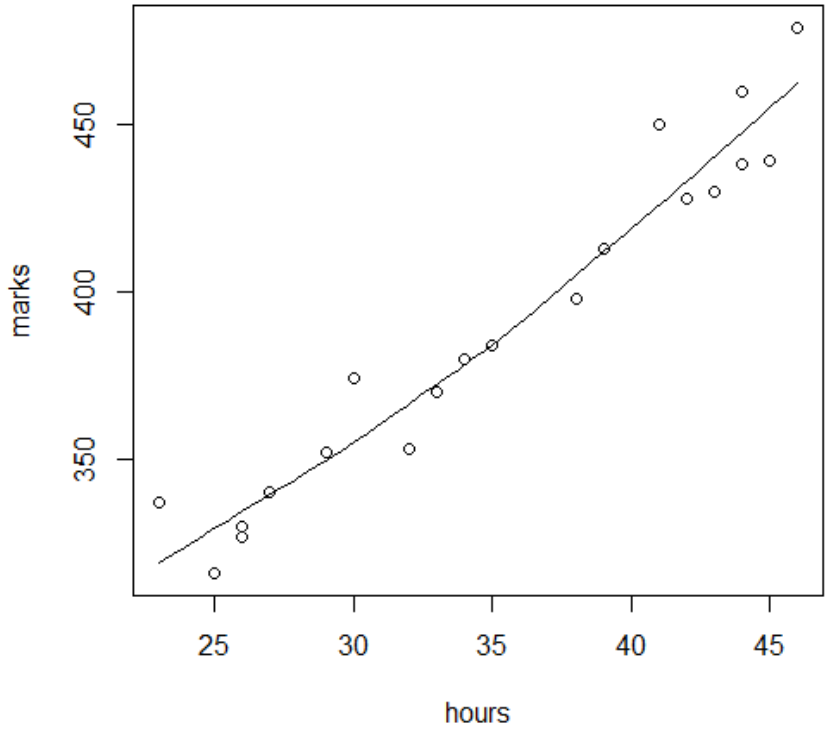
```
marks =
c(337,316,327,340,374,330,352,353,370,380,384,39
8,413,428,430,438,439,479,460,450)

hours =
c(23,25,26,27,30,26,29,32,33,34,35,38,39,42,43,4
4,45,46,44,41)
```

Scatter Plots with Smooth Curve

Example

`scatter.smooth(x,y)` provides scatter plot with smooth curve

Example: `scatter.smooth(hours, marks)`



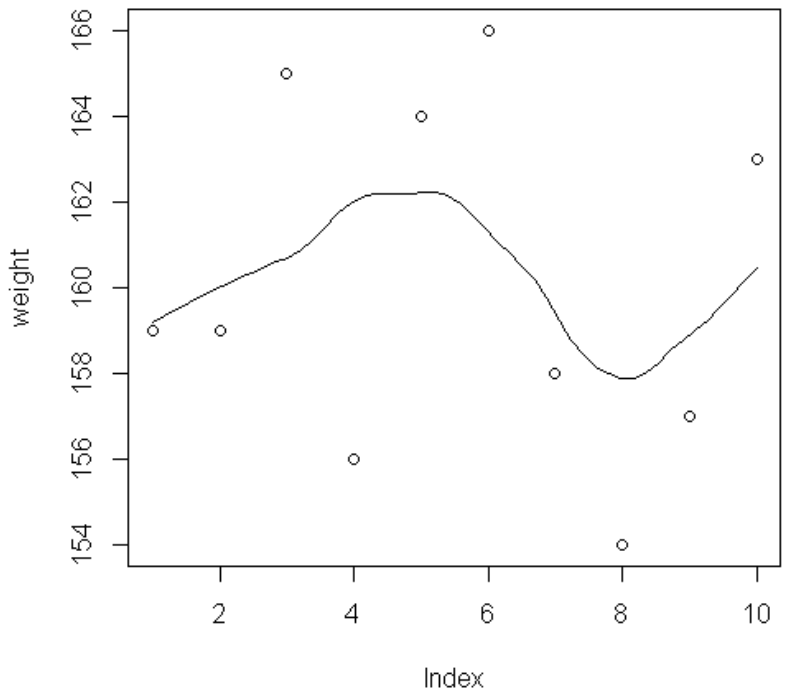
Scatter Plots with Smooth Curve

Example: Single Variable

Weights of 10 bags of grain are obtained (in kg.) and recorded as follows:

159 159 165 156 164 166 158 154 157 163

```
> weight = c(159, 159, 165, 156, 164, 166, 158, 154, 157, 163)
> scatter.smooth(weight)
```



Scatter Plots with Smooth Curve

Example: Single Variable

Height of 50 persons are recorded (in Centimetres) as follow:

166,125,130,142,147,159,159,147,165,156,149,164,137,166,135,142,

133,136,127,143,165,121,142,148,158,146,154,157,124,125,158,159,

164,143,154,152,141,164,131,152,152,161,143,143,139,131,125,145,

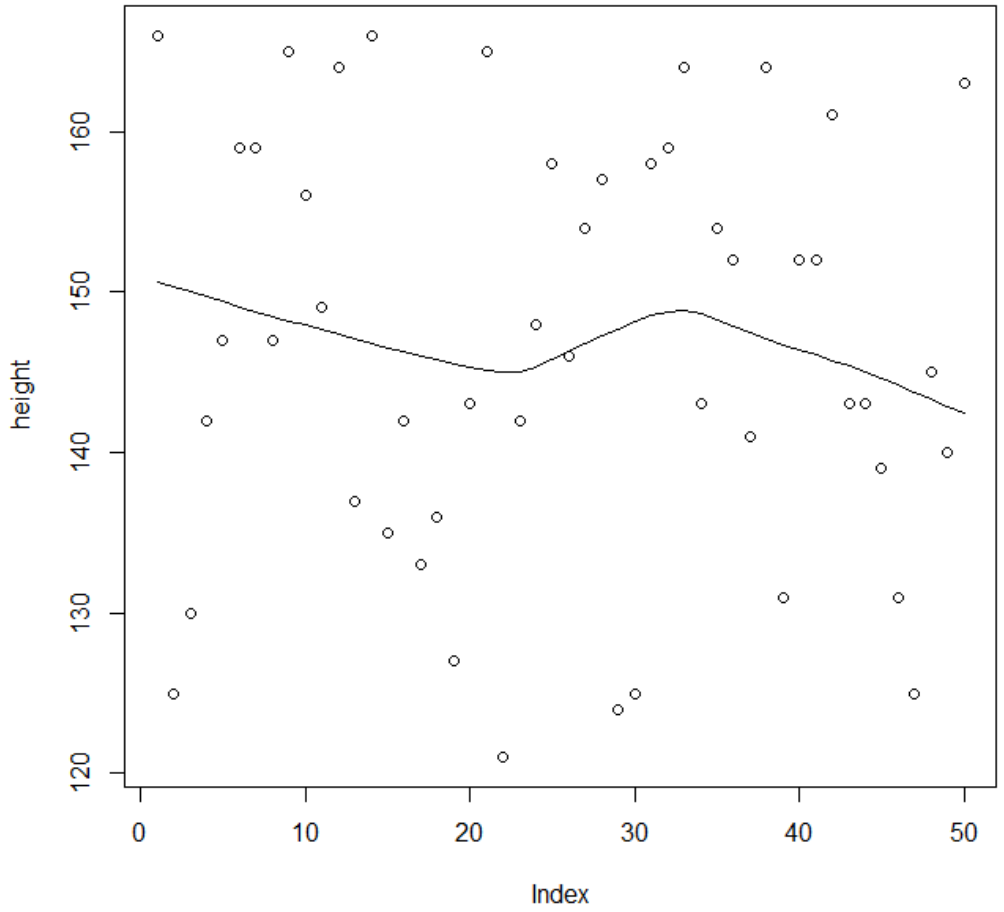
140,163

```
> height = c(166,125,130,142,147,159,159,147,
165,156,149,164,137,166,135,142,133,136,127,143,
165,121,142,148,158,146,154,157,124,125,158,159,
164,143,154,152,141,164,131,152,152,161,143,143,
139,131,125,145,140,163)
```

Scatter Plots with Smooth Curve

Example: Single Variable

```
> scatter.smooth(height)
```



Smooth Scatter Plot

Other options are available.

```
scatter.smooth(x, y = NULL, span = 2/3, degree =
1, family = c("symmetric", "gaussian"), xlab =
NULL, ylab = NULL, ylim = range(y, pred$y, na.rm
= TRUE), evaluation = 50, ..., lpars = list())
```

Smooth Scatter Plots

smoothScatter produces a smoothed colour density representation of a scatterplot, obtained through a (2D) kernel density estimate.

```
smoothScatter(x)
```

```
smoothScatter(x, y = NULL, nbin = 128, bandwidth
colramp = colorRampPalette(c("white",blues9)),
              nrpoints = 100, ret.selection = FALSE,
              pch = ".", cex = 1, col = "black",
              transformation = function(x) x^.25,
              postPlotHook = box, xlab = NULL,
              ylab = NULL, xlim, ylim, xaxs =
              par("xaxs"), yaxs = par("yaxs"), ...)
```

Smooth Scatter Plots

`x, y`	`x` and `y` arguments provide the x and y coordinates for the plot. If supplied separately, they must be of the same length.
`nbin`	numeric vector of length one (for both directions) or two (for x and y separately) specifying the number of equally spaced grid points for the density estimation.
`bandwidth`	numeric vector (length 1 or 2) of smoothing bandwidth
`colramp`	function accepting an integer n as an argument and returning n colours.
`nrpoints`	number of points to be superimposed on the density image. The first `nrpoints` points from those areas of lowest regional densities will be plotted.
`ret.selection`	logical indicating to return the ordered indices of "low density" points if nrpoints > 0.

Smooth Scatter Plots

`pch, cex, col`	arguments passed to points, when nrpoints > 0: point symbol, character expansion factor and colour.
`transformation`	function mapping the density scale to the colour scale.
`postPlotHook`	either NULL or a function which will be called (with no arguments) after image.
`xlab, ylab`	character strings to be used as axis labels, passed to image.
`xlim, ylim`	numeric vectors of length 2 specifying axis limits.

Use `help("smoothScatter")` to get more details.

Smooth Scatter Plots

Example: Single Variable

Weights of 10 bags of grain are obtained (in Kg.) and recorded as follows:

159 159 165 156 164 166 158 154 157 163

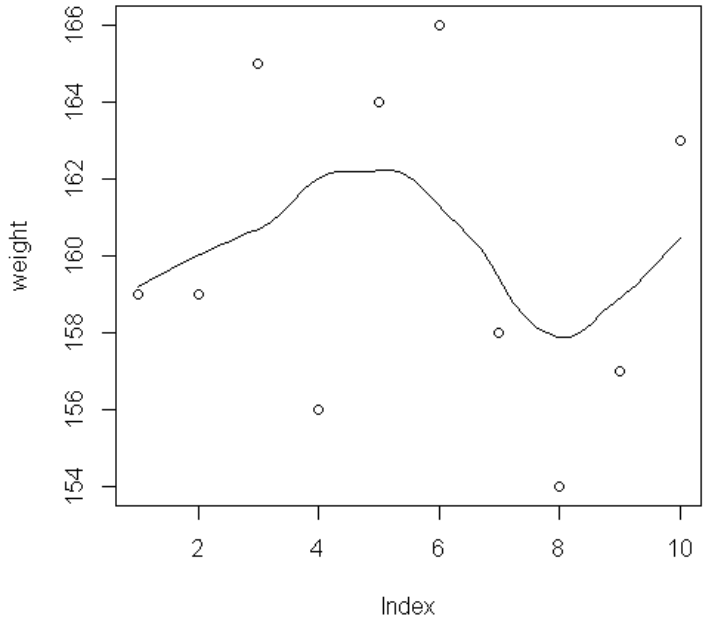
```
> weight = c(159, 159, 165, 156, 164, 166, 158, 154, 157, 163)
```

We had earlier obtained the

smooth scatter plot with curve

as follows by using the command

```
scatter.smooth(weight)
```

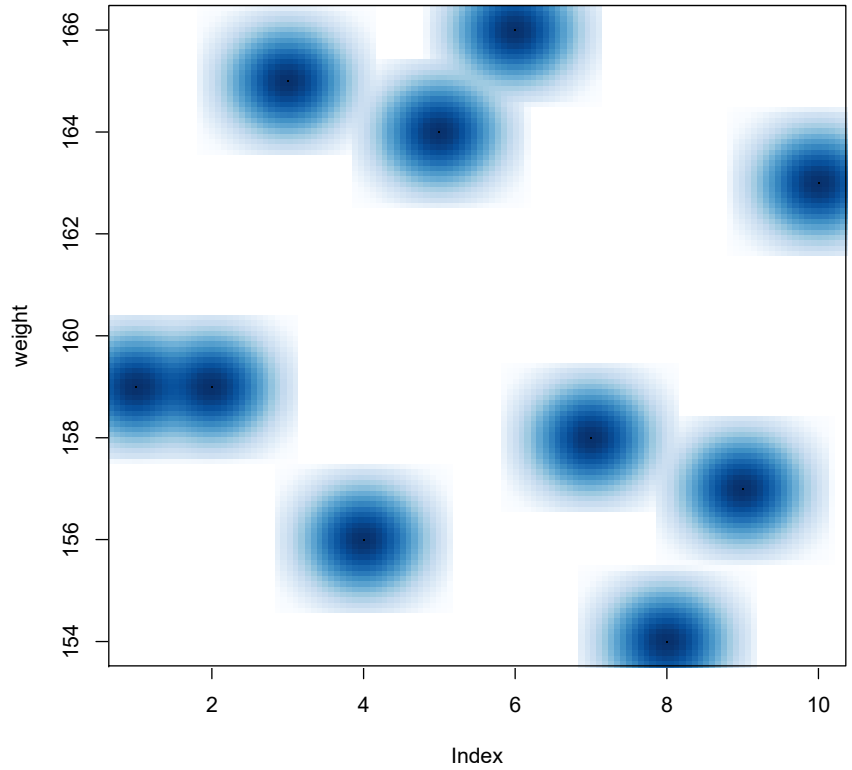


Now we obtain smooth scatter plot.

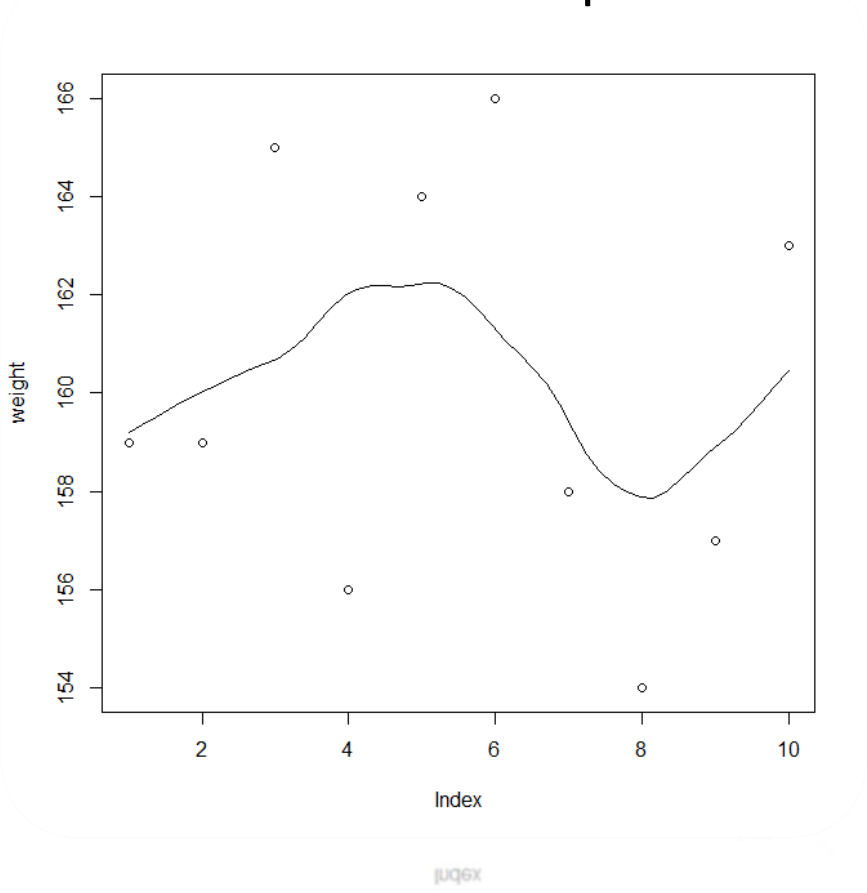
Smooth Scatter Plots

Example

```
> weight <-c(159, 159, 165, 156, 164, 166,
158, 154, 157, 163)
> smoothScatter(weight)
```



Recall the scatter plot



Smooth Scatter Plots

Example

Height of 50 persons are recorded (in Centimetres) as follow:

166,125,130,142,147,159,159,147,165,156,149,164,137,166,135,142,

133,136,127,143,165,121,142,148,158,146,154,157,124,125,158,159,

164,143,154,152,141,164,131,152,152,161,143,143,139,131,125,145,

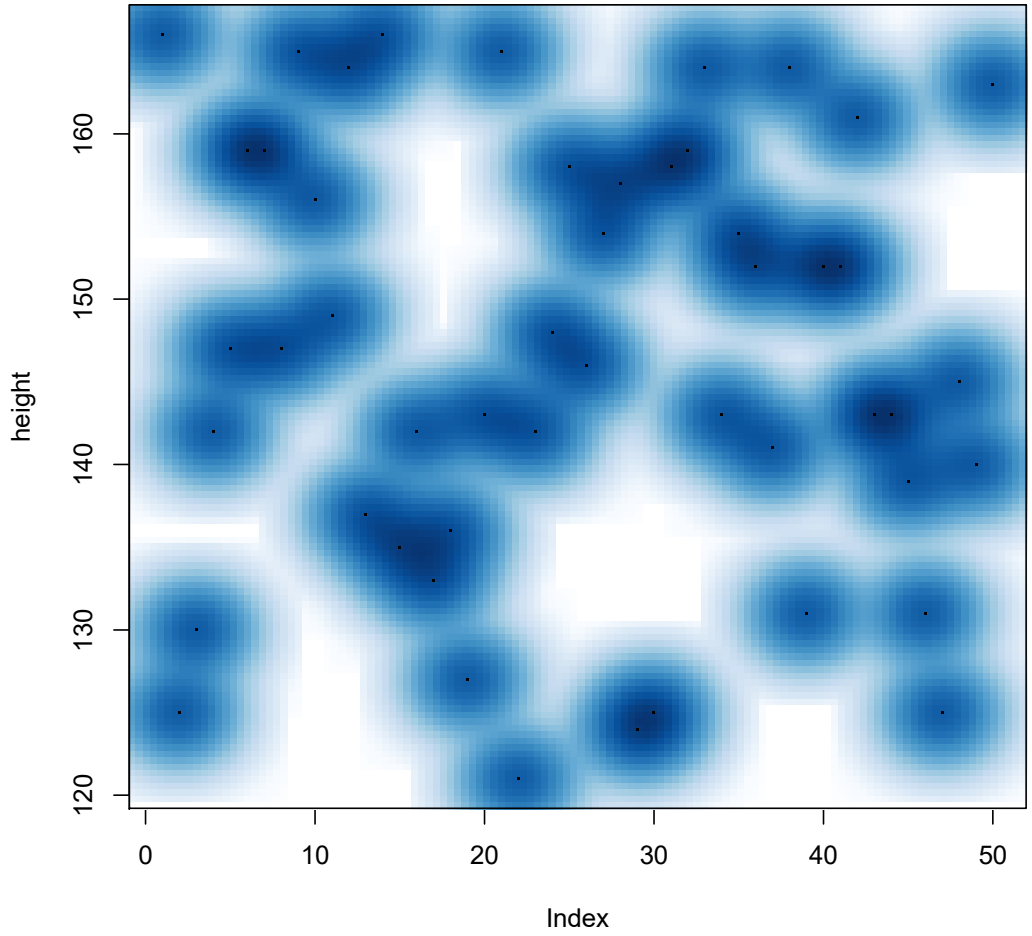
140,163

```
> height = c(166,125,130,142,147,159,159,147,
165,156,149,164,137,166,135,142,133,136,127,143,
165,121,142,148,158,146,154,157,124,125,158,159,
164,143,154,152,141,164,131,152,152,161,143,143,
139,131,125,145,140,163)
```

Smooth Scatter Plots

Example

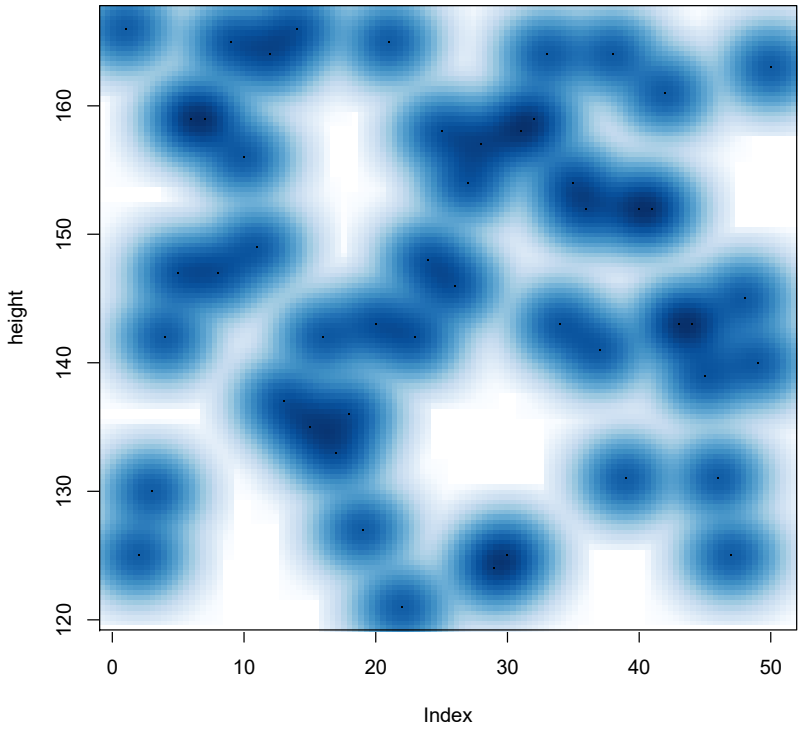
```
> smoothScatter(height)
```



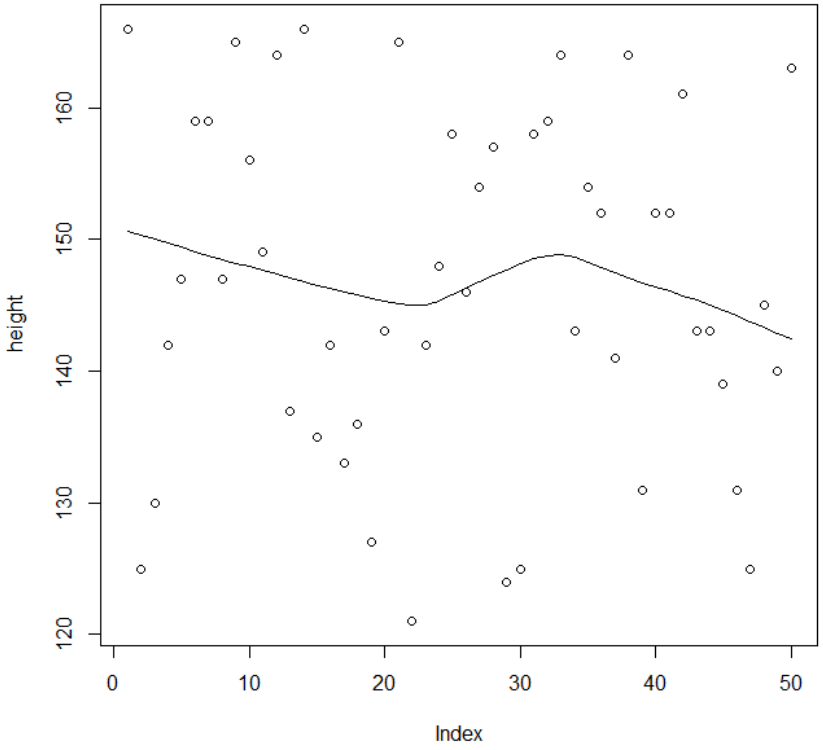
Smooth Scatter Plots

Example

```
> smoothScatter(height)
```



Recall the scatter plot



Scatter Plots with Smooth Curve

Example

```
marks =
c(337,316,327,340,374,330,352,353,370,380,384,39
8,413,428,430,438,439,479,460,450)

hours =
c(23,25,26,27,30,26,29,32,33,34,35,38,39,42,43,4
4,45,46,44,41)

smoothScatter(hours, marks)
```

Scatter Plots with Smooth Curve

Example

```
smoothScatter(hours, marks)
```