

Section 5: Bivariate plots for two continuous variables

Review of important functions covered in this section:

Functions

plot	Base plot function
scatterplot	Scatterplot
boxplot2D	Boxplot of projection of 2D data (aplpack package)
bagplot	Bivariate boxplot (aplpack package)
pairs	Scatterplot matrices
scatterplotMatrix	Scatterplot matrices (car package)

Supplementary functions:

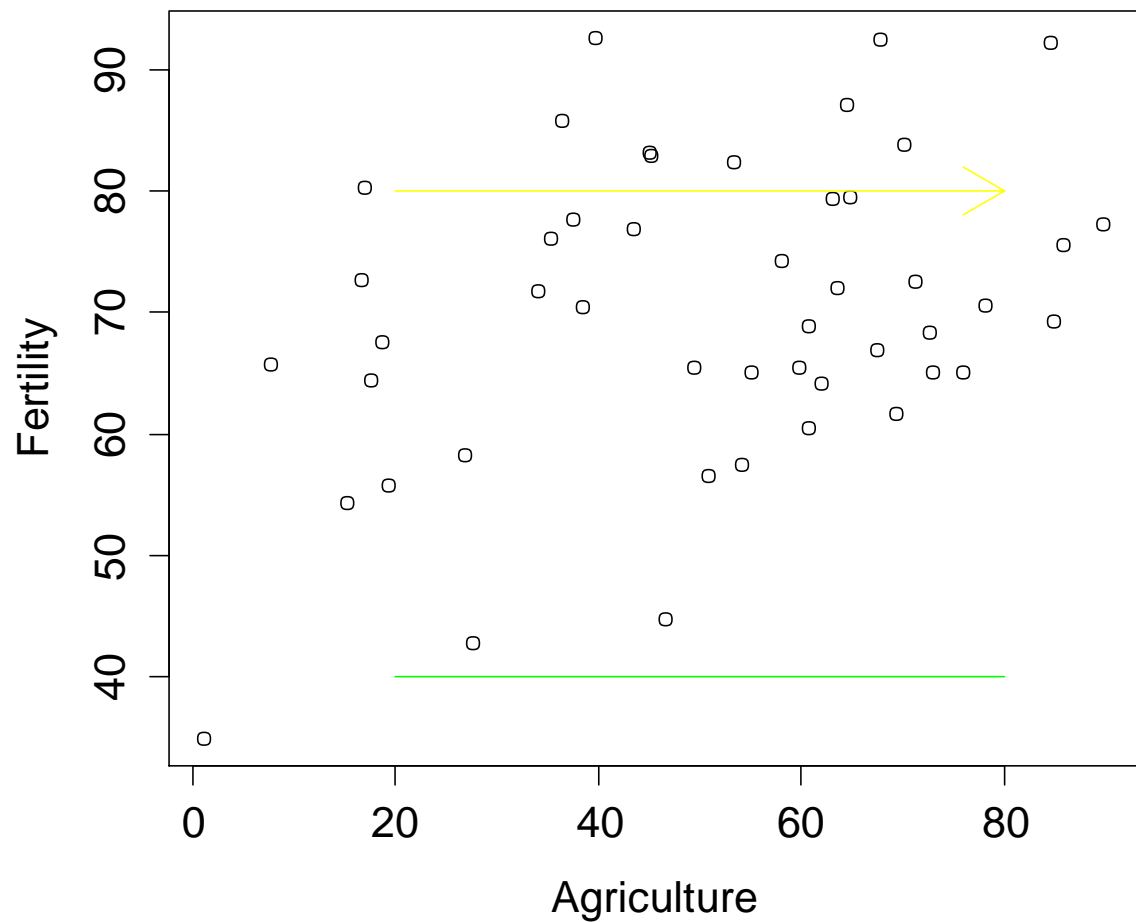
jitter	Adds noise to numbers
abline	Adds a straight line in the existing plot
segments	Adds a line segment in the existing plot
arrows	Adds an arrow line in the existing plot
sample	Generates random sample
bquote	Puts partial substitution in expressions
paste	Concatenates strings
scatter.smooth	Scatterplot with smooth curve (fitted by loess function)
lowess	Locally-weighted polynomial regression smoothing
lm	Fits linear model
residuals	Extracts residuals from a fitted lm object
predict	Generates model predictions using a fitted lm object

Exercise

We'll use `swiss` dataset.

- a. Draw a scatterplot of Agriculture in the horizontal axis and Fertilizer in the vertical axis. Add a green line from 20-80 parallel to the horizontal axis at the vertical axis point 40. Similarly add an arrow with same size at the vertical axis point 80.

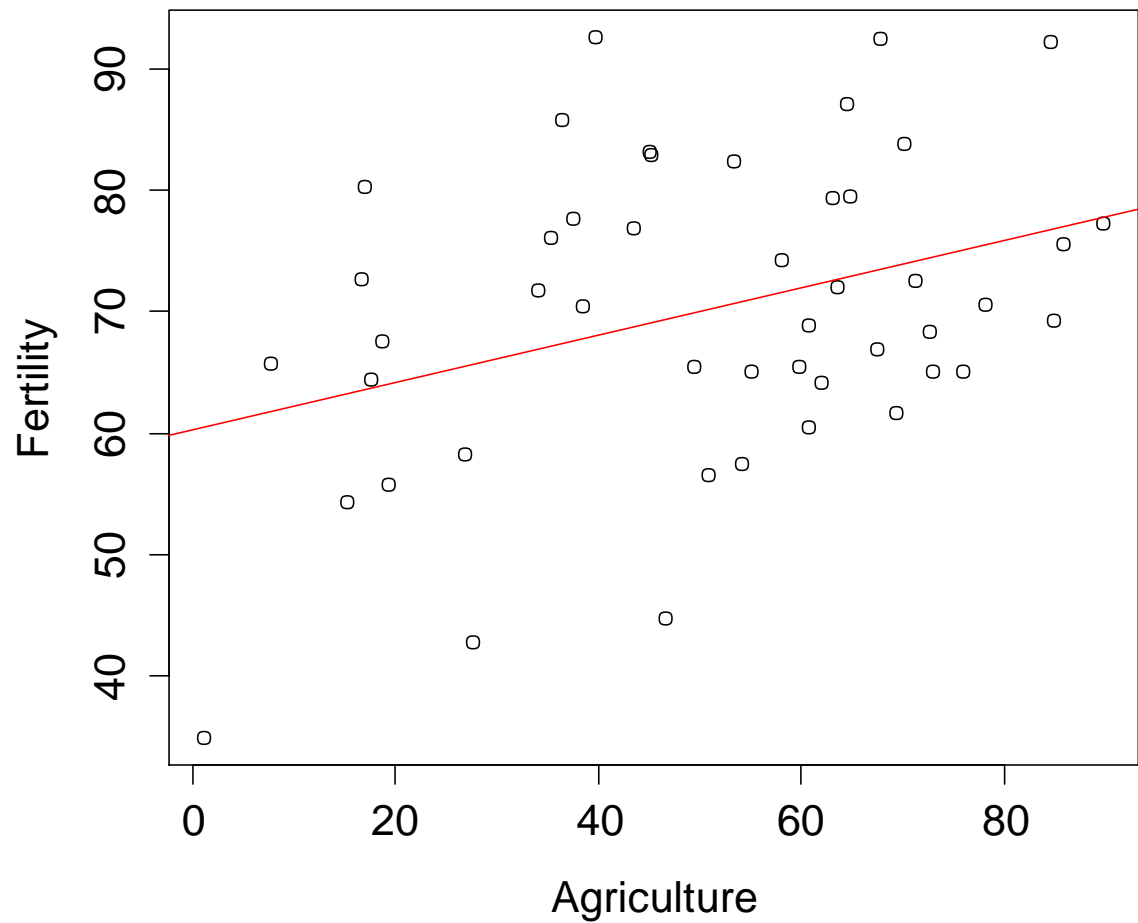
```
with(swiss, plot(Agriculture, Fertility))  
segments(x0=20, y0=40, x1=80, y1=40, col = "green")  
arrows(x0=20, y0=80, x1=80, y1=80, col = "yellow")
```



- b. Draw a scatterplot of Agriculture in the horizontal axis and Fertilizer in the vertical axis using plot function. Add a red regression line in the same plot.

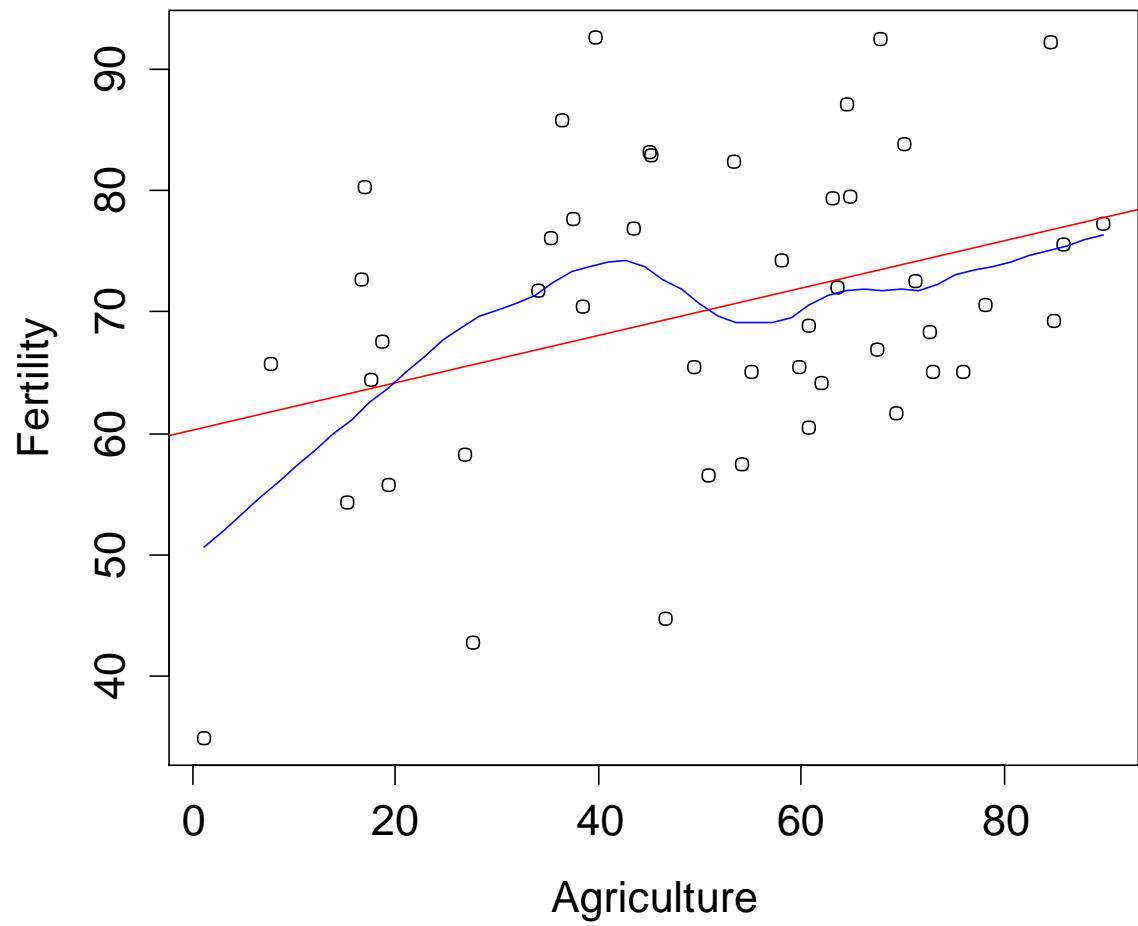
```
with(swiss, plot(Agriculture, Fertility))
```

```
abline(lm(Fertility ~ Agriculture, data = swiss), col = "red")
```



c. Add a blue loess fit in the same plot.

```
with(swiss, lines(loess.smooth(Agriculture, Fertility,  
span=.5), col = "blue"))
```

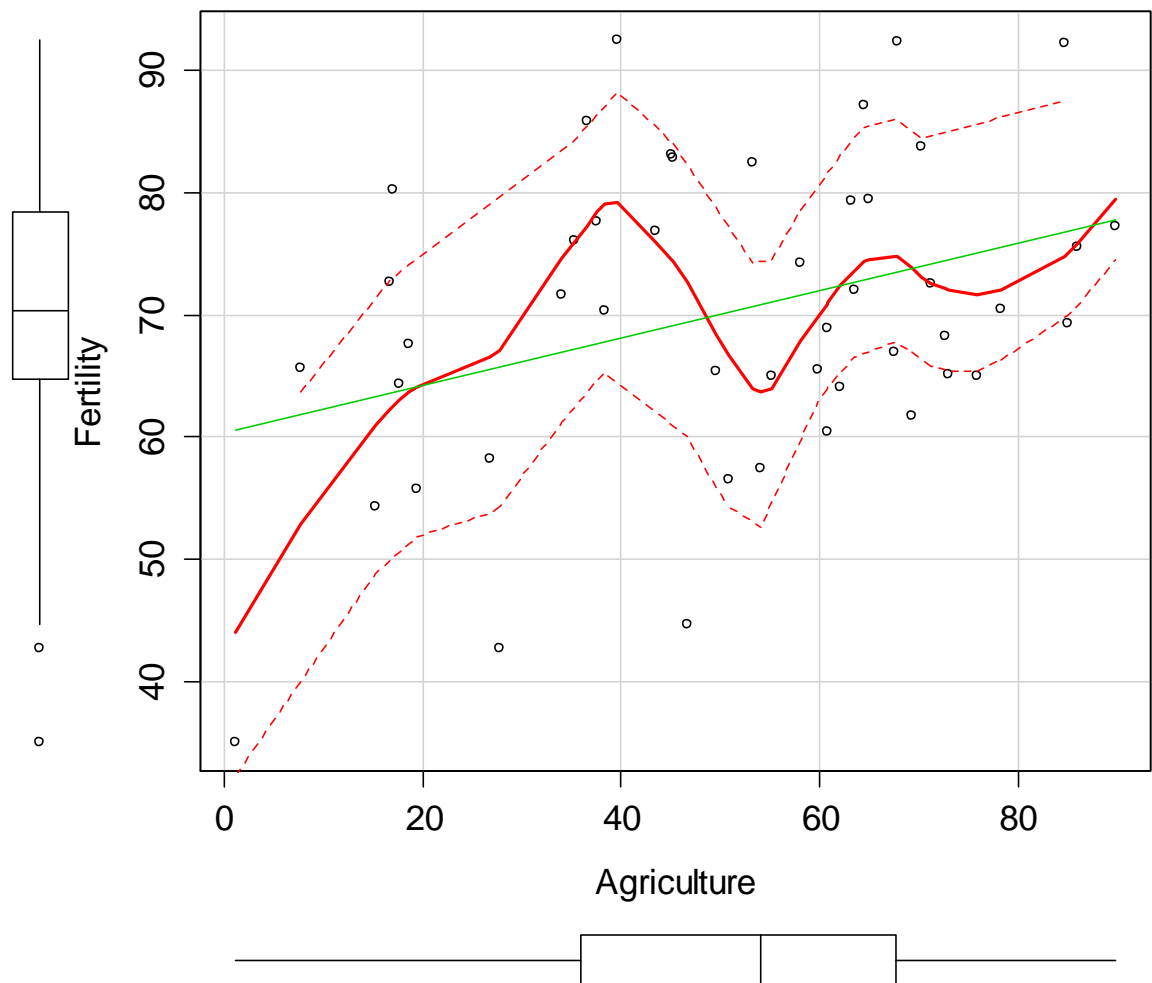


- d. Install and load the package “car”. Use this package to draw scatterplot of Agriculture in the horizontal axis and Fertility in the vertical axis.

```
install.packages("car")
```

```
require("car")
```

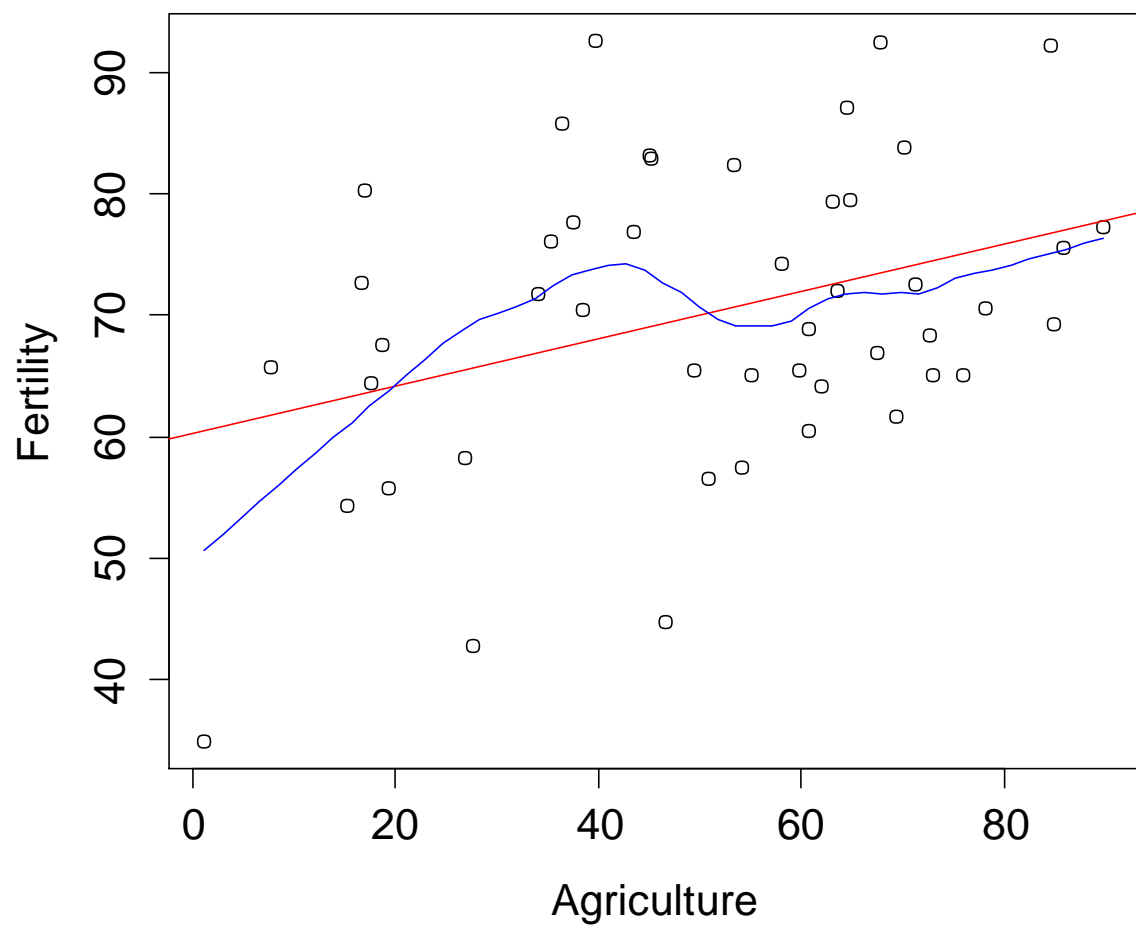
```
with(swiss, scatterplot(Agriculture, Fertility))
```



- e. Draw scatterplot of Agriculture in the horizontal axis and Fertility in the vertical axis. Add regression and loess line in it. Find **correlation** between these two variables (0.35) using `cor` function in two **significant digits**. Report this correlation as a title of the plot in the following format: $\hat{\rho} = 0.35$.

```
with(swiss, plot(Agriculture, Fertility))
abline(lm(Fertility ~ Agriculture, data = swiss), col = "red")
with(swiss, lines(loess.smooth(Agriculture, Fertility,
span=.5), col = "blue"))
cor1 = round(with(swiss, cor(Fertility, Agriculture)),2)
title(main = bquote(paste(hat(rho), " = ", .(cor1))))
```

$$\hat{\rho} = 0.35$$



- f. Make a layout of the plot window to accommodate exactly 5 plots simultaneously. Draw five scatterplots where Fertility is in the vertical axis and rest of the variables in the dataset (Agriculture, Examination, Education, Catholic, Infant.Mortality) in the horizontal axes. Add corresponding regression lines, loess lines and correlation as a title as instructed in the previous question.

```
layout(matrix(c(1,2,3,4,5,5), byrow = T, nrow = 3))
```

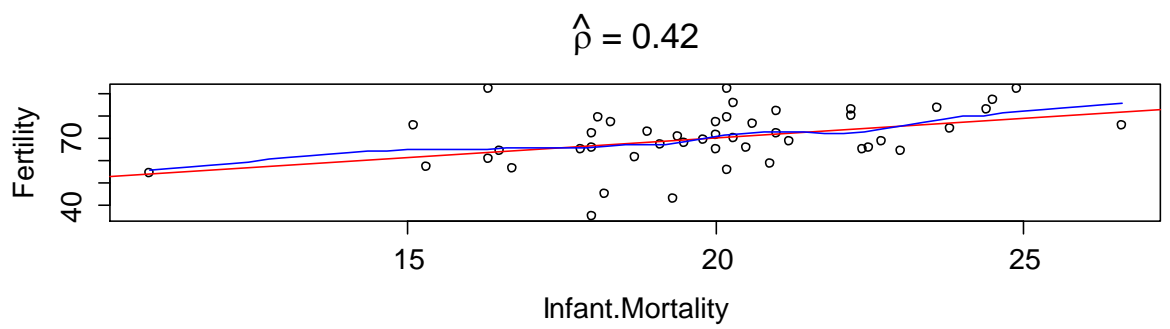
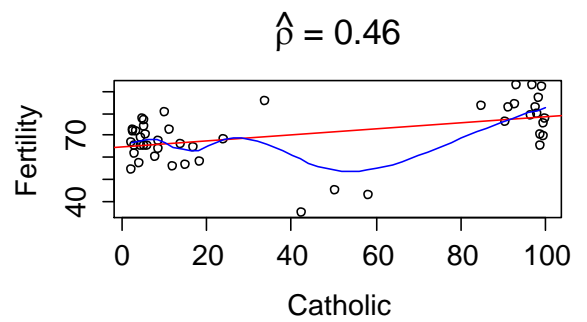
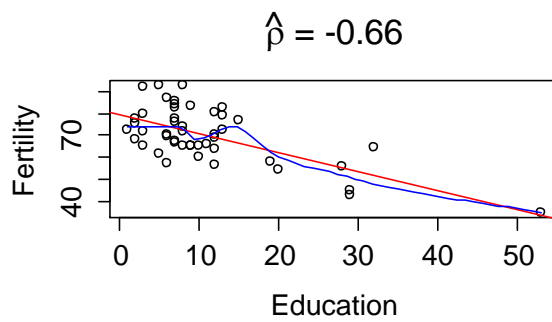
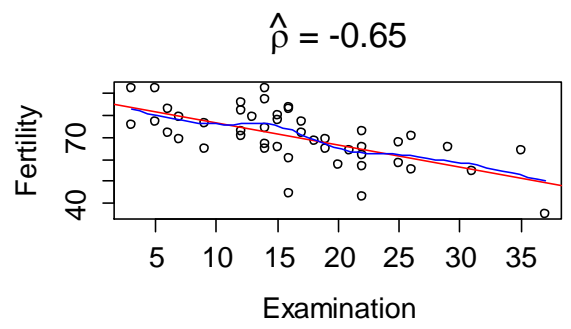
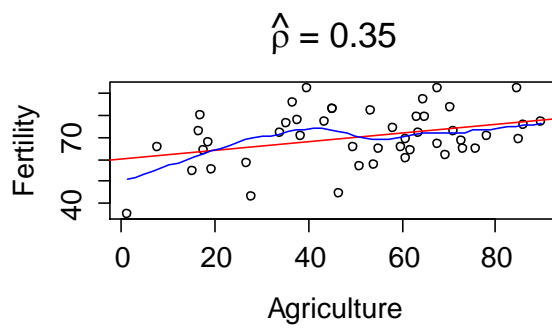
```
with(swiss, plot(Agriculture, Fertility))
abline(lm(Fertility ~ Agriculture, data = swiss), col = "red")
with(swiss, lines(loess.smooth(Agriculture, Fertility,
span=.5), col = "blue"))
cor1 = round(with(swiss, cor(Fertility, Agriculture)),2)
title(main = bquote(paste(hat(rho), " = ", .(cor1))))
```

```
with(swiss, plot(Examination, Fertility))
abline(lm(Fertility ~ Examination, data = swiss), col = "red")
with(swiss, lines(loess.smooth(Examination, Fertility,
span=.5), col = "blue"))
cor1 = round(with(swiss, cor(Fertility, Examination)),2)
title(main = bquote(paste(hat(rho), " = ", .(cor1))))
```

```
with(swiss, plot(Education, Fertility))
abline(lm(Fertility ~ Education, data = swiss), col = "red")
with(swiss, lines(loess.smooth(Education, Fertility, span=.5),
col = "blue"))
cor1 = round(with(swiss, cor(Fertility, Education)),2)
title(main = bquote(paste(hat(rho), " = ", .(cor1))))
```

```
with(swiss, plot(Catholic, Fertility))
abline(lm(Fertility ~ Catholic, data = swiss), col = "red")
with(swiss, lines(loess.smooth(Catholic, Fertility, span=.5),
col = "blue"))
cor1 = round(with(swiss, cor(Fertility, Catholic)),2)
title(main = bquote(paste(hat(rho), " = ", .(cor1))))
```

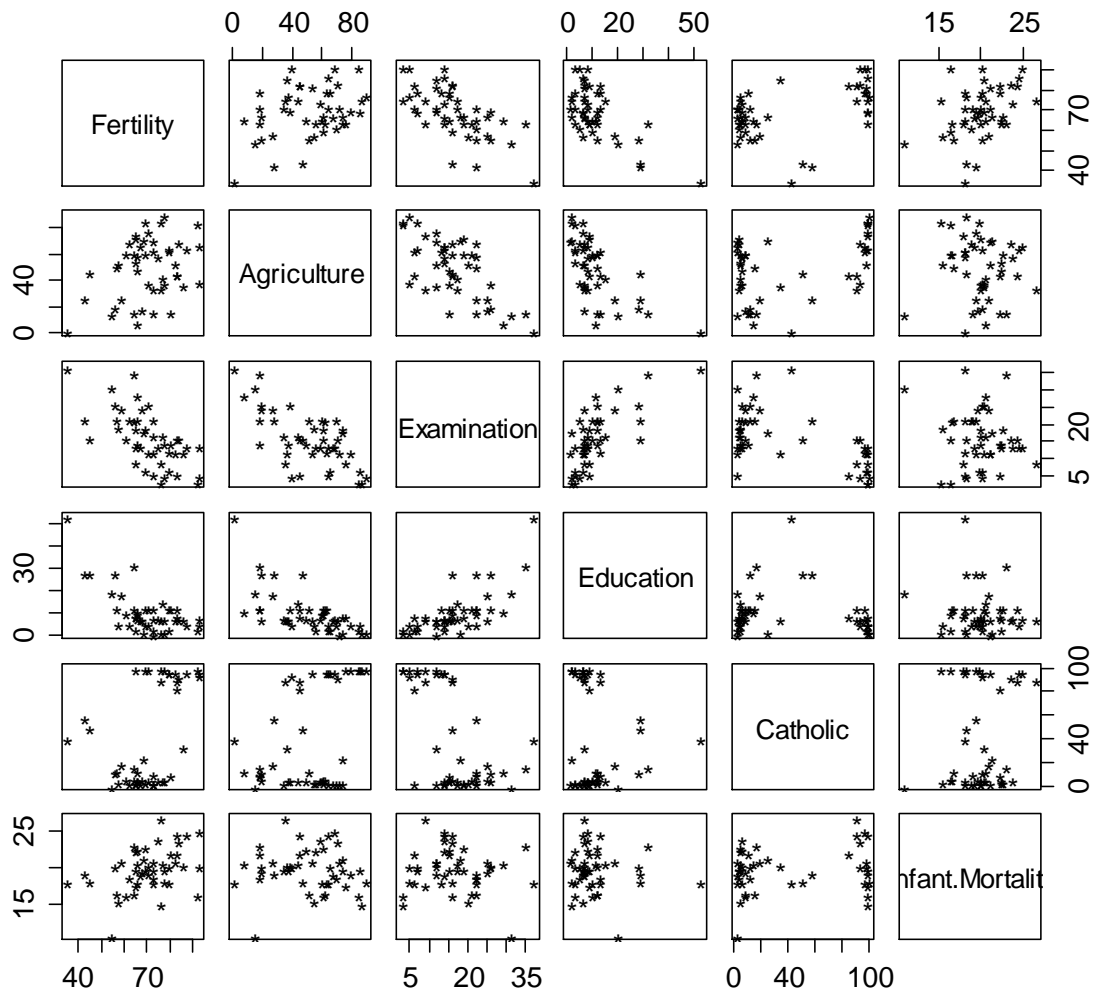
```
with(swiss, plot(Infant.Mortality, Fertility))
abline(lm(Fertility ~ Infant.Mortality, data = swiss), col =
"red")
with(swiss, lines(loess.smooth(Infant.Mortality, Fertility,
span=.5), col = "blue"))
cor1 = round(with(swiss, cor(Fertility, Infant.Mortality)),2)
title(main = bquote(paste(hat(rho), " = ", .(cor1))))
```



- g. Bring back default plot layout. Draw a scatterplot matrix of all the variables in the dataset where each scatterplot points will be represented by a *.

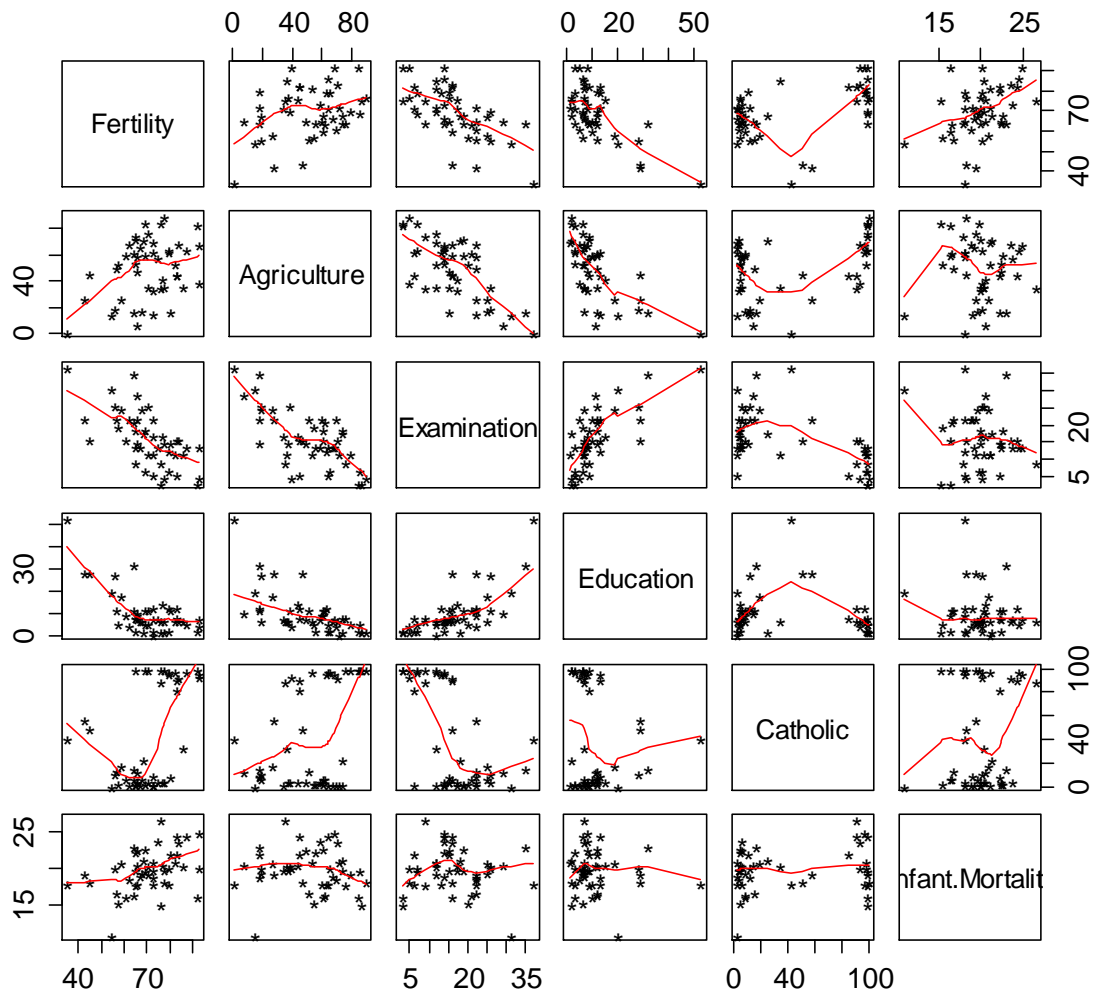
```
layout(1)
```

```
pairs(swiss, pch = "*")
```



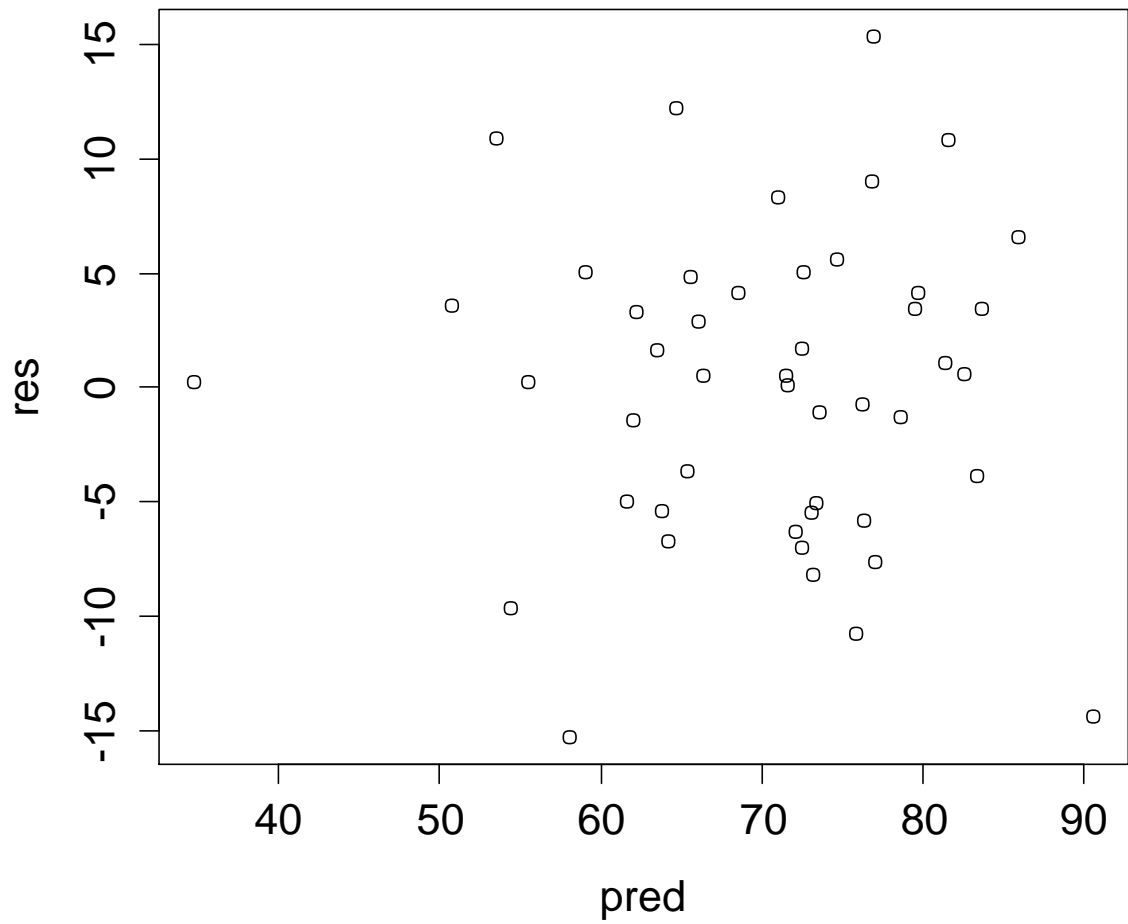
h. Add smooth lines in each cell of the scatterplot matrix.

```
pairs(swiss, pch = "*", panel = panel.smooth)
```



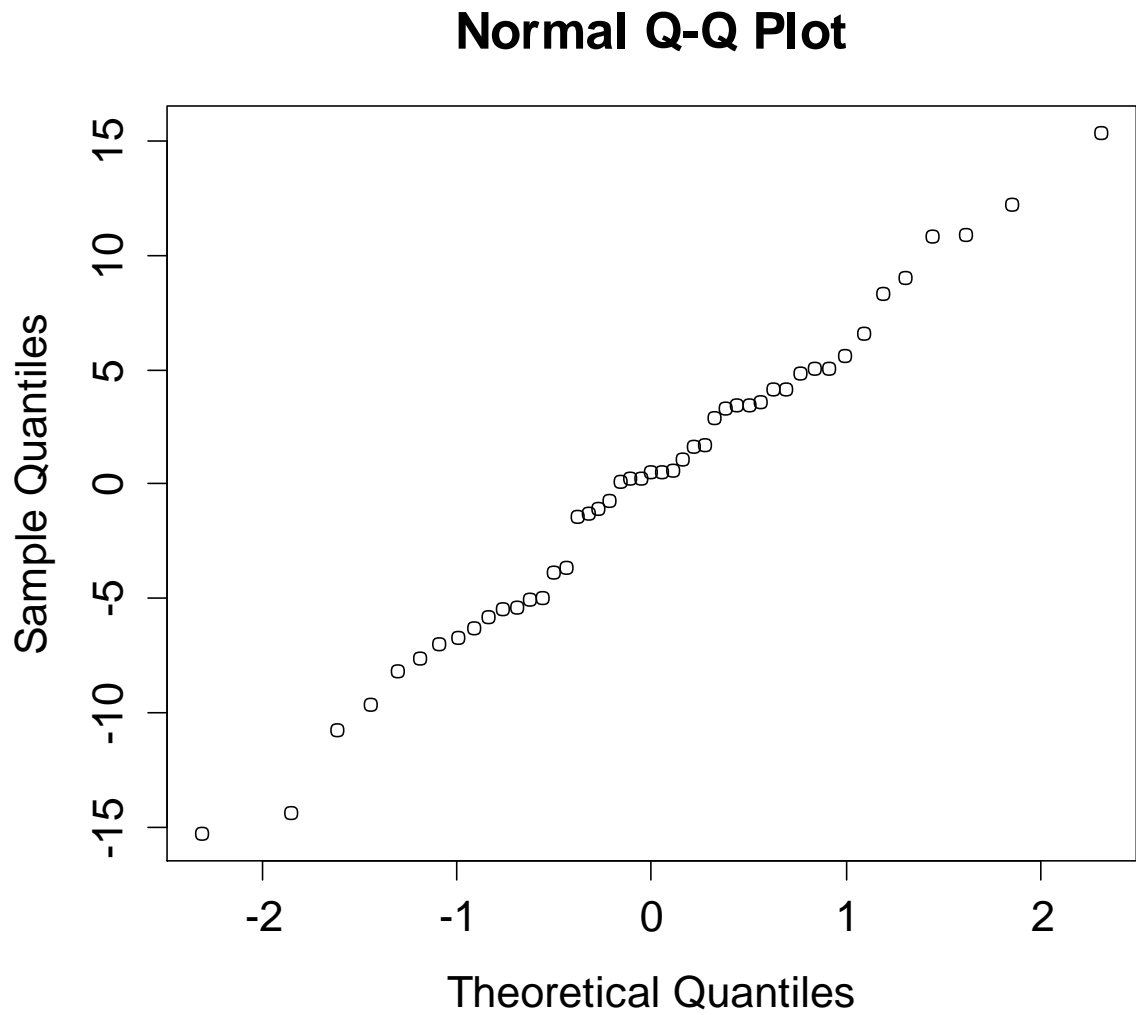
- i. Fit a regression line with Fertility being dependent variable and rest being predictor variables. Draw a plot of residual versus fitted values.

```
fit = lm(Fertility ~ ., data = swiss)
pred = predict(fit)
res = residuals(fit)
plot(pred, res)
```



j. Draw a quantile-quantile plot of the residuals.

```
qqnorm(res)
```



k. Plot all the default regression diagnostic plots (4 plots) simultaneously.

```
par(mfrow=c(2,2))
```

```
plot(fit)
```

