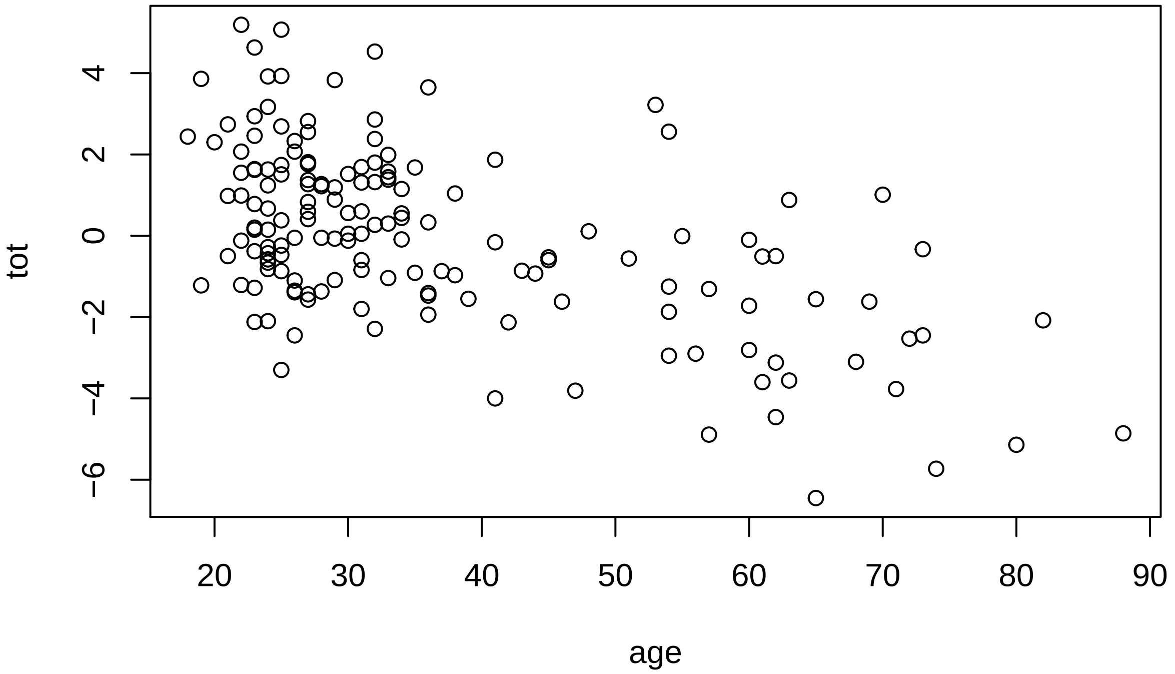
# Case study: kidney data

**From Efron and Hastie (2016), pages 4-8**.

The figure below concerns a study of kidney function,



Data points $(x_i, y_i)$ have been observed for $n = 157$ healthy volunteers, with $x_i$ the $i$th volunteer's age in years, and $y_i$ a composite measure of overall function. Kidney function generally declines with age, as evident in the downward scatter of the points.

The rate of decline is an important question in kidney transplantation: in the past, potential donors past age 60 were prohibited, though, given a shortage of donors, this is no longer enforced.

A potential new donor, aged 55, has appeared, and we wish to assess his kidney fitness without subjecting him to an arduous series of medical tests.

1. Download the data

```
kidney_url <- "https://web.stanford.edu/~hastie/CASI_files/DATA/kidney.txt"
kidney <- read.table(kidney_url, header=TRUE, sep=" ")
```

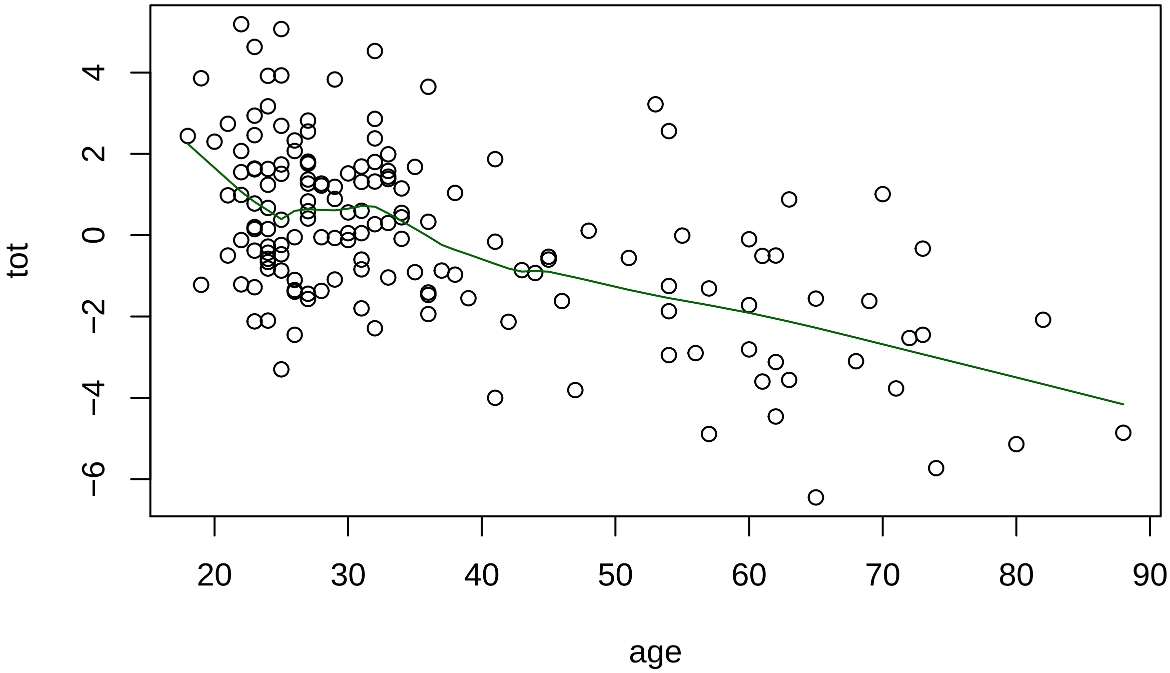
2. Fit the regression line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ and compute the estimates of kidney fitness and their standard errors for ages $= 20, 30, 40, 50, 60, 70, 80$.

```
fit <- lm(tot ~ age, data=kidney)
ages <- data.frame(age = seq(20, 80, 10))
res <- predict(fit, ages, se.fit = TRUE)
print(round(data.frame(age=ages$age,estimate=res$fit,se=res$se.fit),2))
```

```
##   age estimate   se
## 1  20     1.29 0.21
## 2  30     0.50 0.15
## 3  40    -0.28 0.15
## 4  50    -1.07 0.19
## 5  60    -1.86 0.26
## 6  70    -2.64 0.34
## 7  80    -3.43 0.42
```

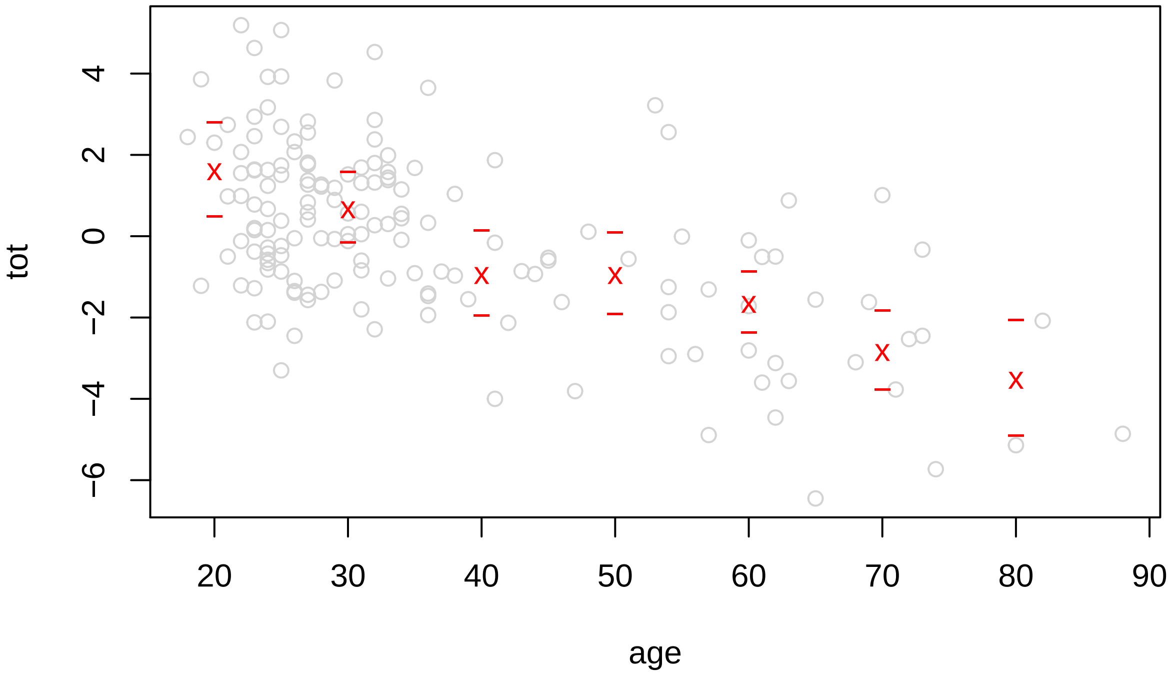
3. Fit a local polynomial by using the function `lowess(x,y,1/3)`

```
plot(kidney)
lines(lowess(kidney, f = 1/3), col="darkgreen")
```



4. Fit a regression spline by using a natural cubic spline with 10 degrees of freedom. Plot the eestimates of kidney fitness for ages $= 20, 30, 40, 50, 60, 70, 80$ along with 95% confidence intervals.

```
require(splines)
fit = lm(tot ~ ns(age,10), kidney)
res = predict(fit, newdata=ages, interval = "confidence")
plot(kidney, col="lightgray")
points(ages$age,res[,1], pch="x",col=2)
points(ages$age,res[,2], pch="-",col=2 )
points(ages$age,res[,3], pch="-",col=2 )
```



5. With the model fitted in 4., compute a 95% prediction interval for a new donor of age 55

```
require(splines)
predict(fit, newdata=data.frame(age=55), interval = "prediction")
```

```
##         fit       lwr      upr
## 1 -1.060266 -4.803715 2.683183
```