Assignment 1
K-NN regression

Andreu Meca, Geraldo Gariza, Victor Duque and Jan Leyva
Universitat Politècnica de Catalunya

February 15, 2021

K-NN regression

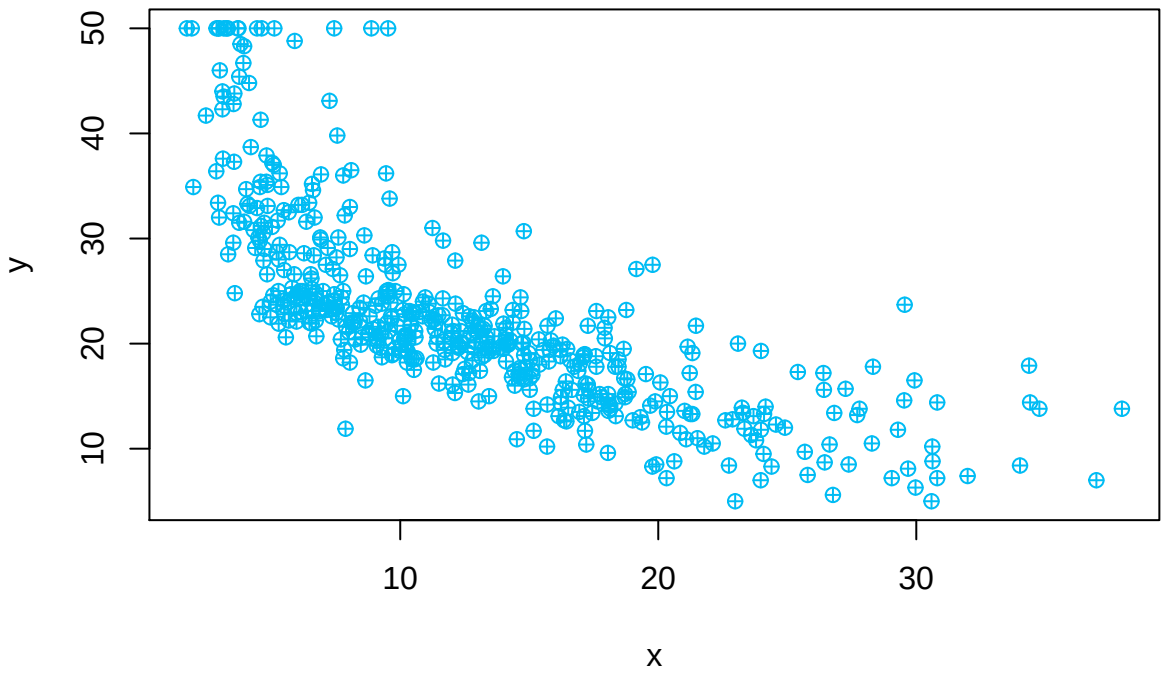
The k nearest-neighbor estimator of $m(t) = \mathrm{E}(\mathrm{Y}|\mathrm{X} = t)$ is defined as

$$\hat{m}(t) = \frac{1}{k} \sum_{i \in N_k(t)} y_i$$

where $\mathrm{N_k}(t)$ is the neighborhood of t defined by the k closest points x_i in the training sample.

Boston housing data from library MASS:

```
library(MASS)
help(Boston)
data(Boston)
x <- Boston$lstat
y <- Boston$medv
plot(x,y, pch = 10,col = "#00BCF4")
```



1. Write a function for computing the k-nn estimator of $m(t)$ for a given value of t $\in R$.

```
k<-5
t<-10
# dist matrix of X's
d_x   <- as.matrix(dist(rbind(t,as.matrix(x))))[1,-1]
d_x_k <- sort(d_x, partial=k)[k]
N_x_k <- unname( which(d_x <= d_x_k) )
(hat_y = sum(y[N_x_k])/k)
```

```
## [1] 21.06
```

```
knn.reg <- function(t, x, y,  k=5){
  d_x   <- as.matrix(dist(rbind(t,as.matrix(x))))[1,-1]
  d_x_k <- sort(d_x, partial=k)[k]
  N_x_k <- unname( which(d_x <= d_x_k) )
  return(hat_y = sum(y[N_x_k])/k)
}
```

Function k-nn for regression code

2. Then, define t as a sequence from 1 to 40: t <- 1:40.

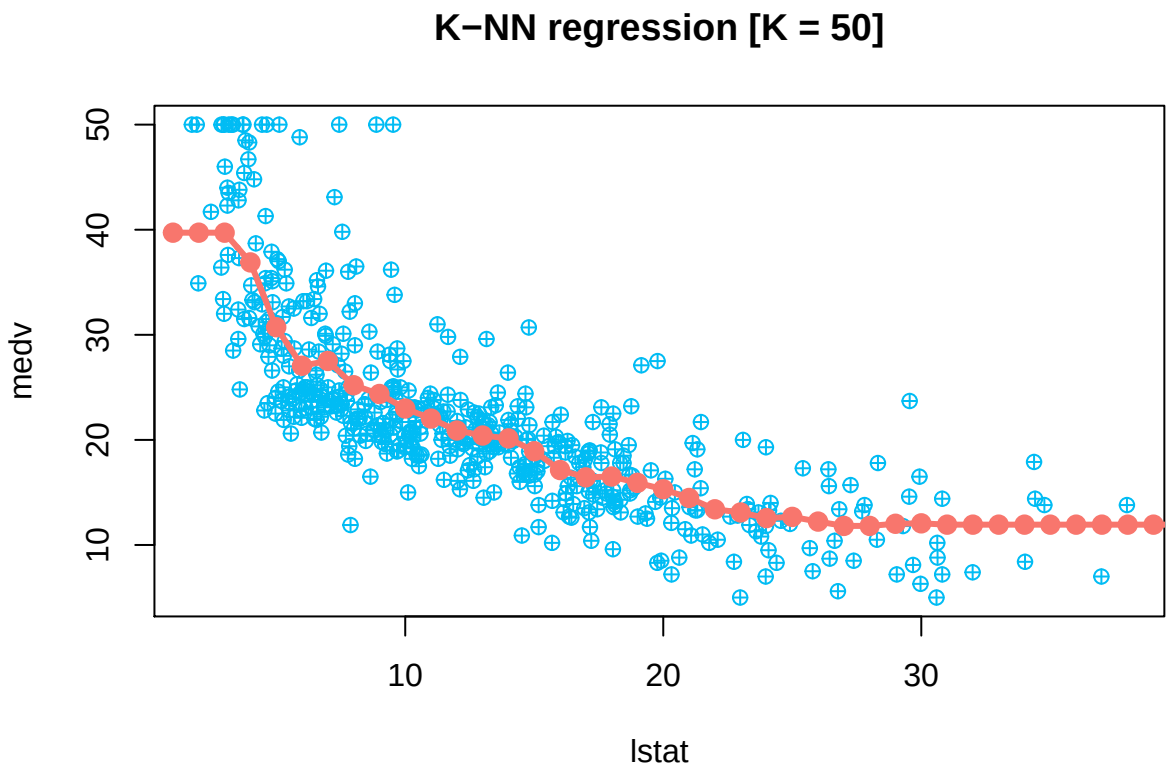
```
t <- 1:40
```

3. Estimate m(t[i]) for i = 1,...,40 using k = 50.

```
nt <- length(t)
hat_p <- matrix(0,1,ncol=nt)

for (i in 1:nt){
    hat_p[,i] <- knn.reg(t[i], x=x, y=y, k = 50)
}
```

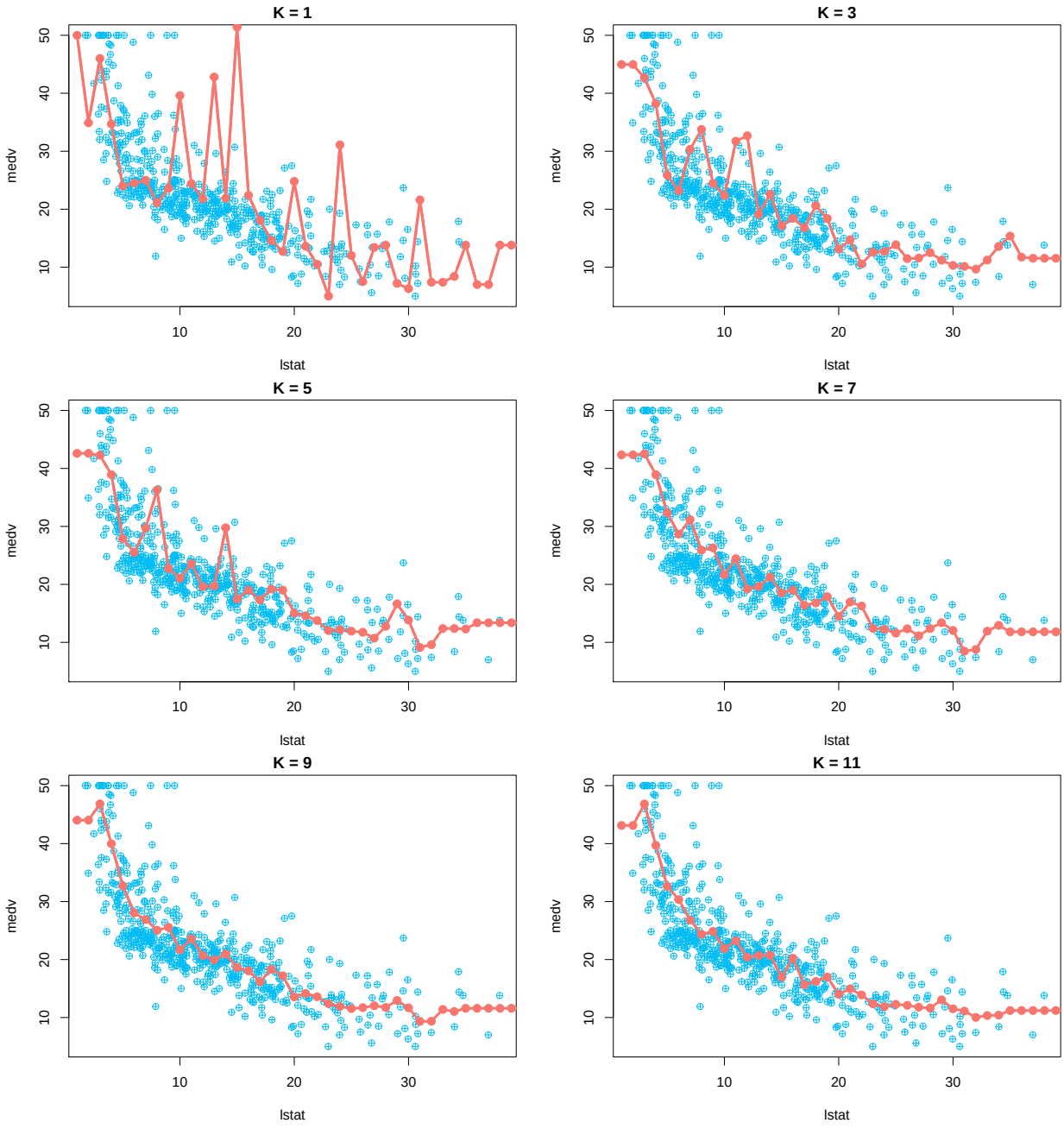
4. Plot y against x. Then represent the estimated regression function.

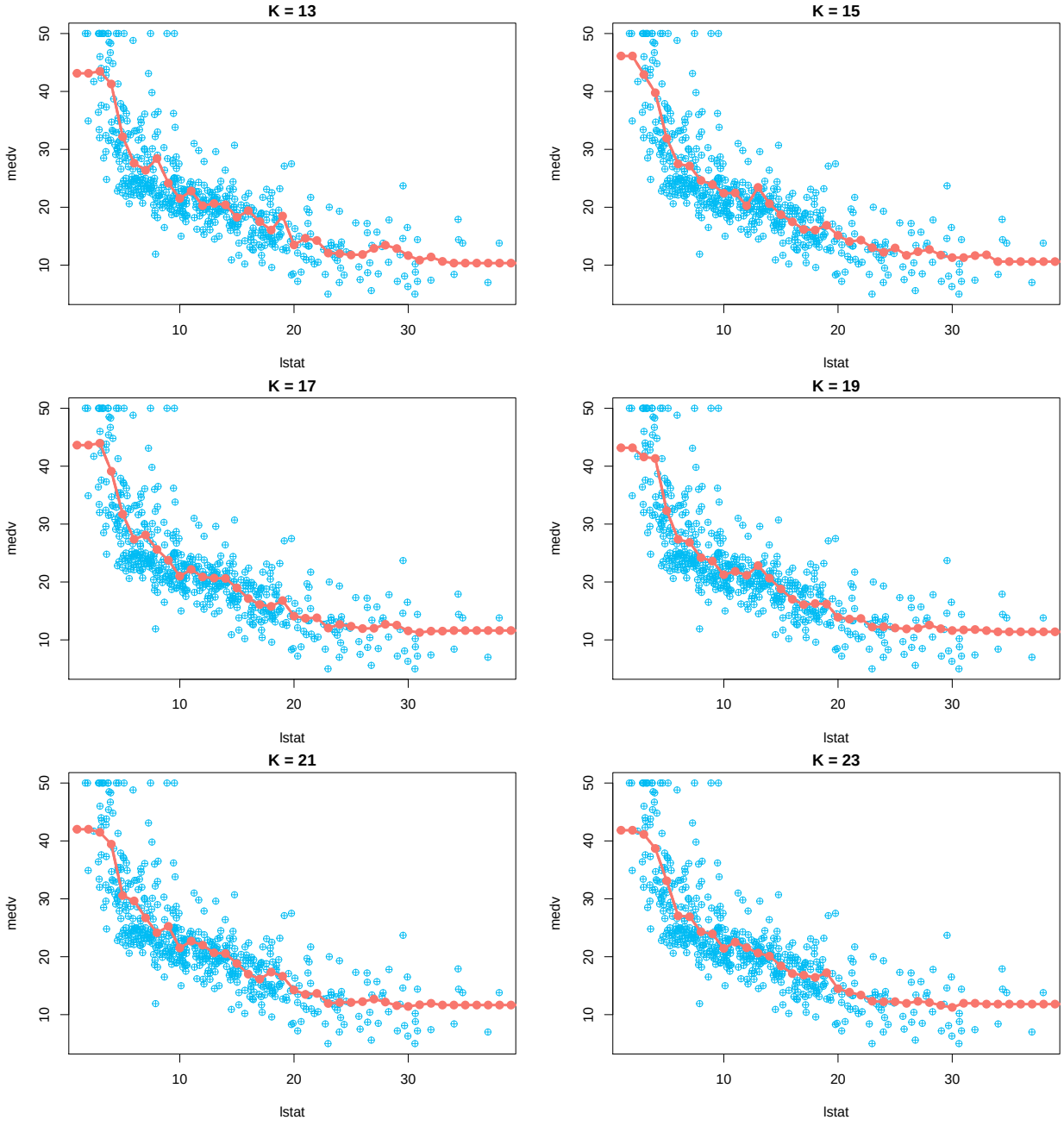
```
plot(x, y, pch = 10,xlab = "lstat", ylab = "medv", col = "#00BCF4")
points(t, hat_p, col="#F8766D" , lwd=3 , pch=19 , type= "l")
points(t, hat_p, col="#F8766D" , lwd=3 , pch=19 , lty = "dashed", type= "b")
title("K-NN regression [K = 50]")
```

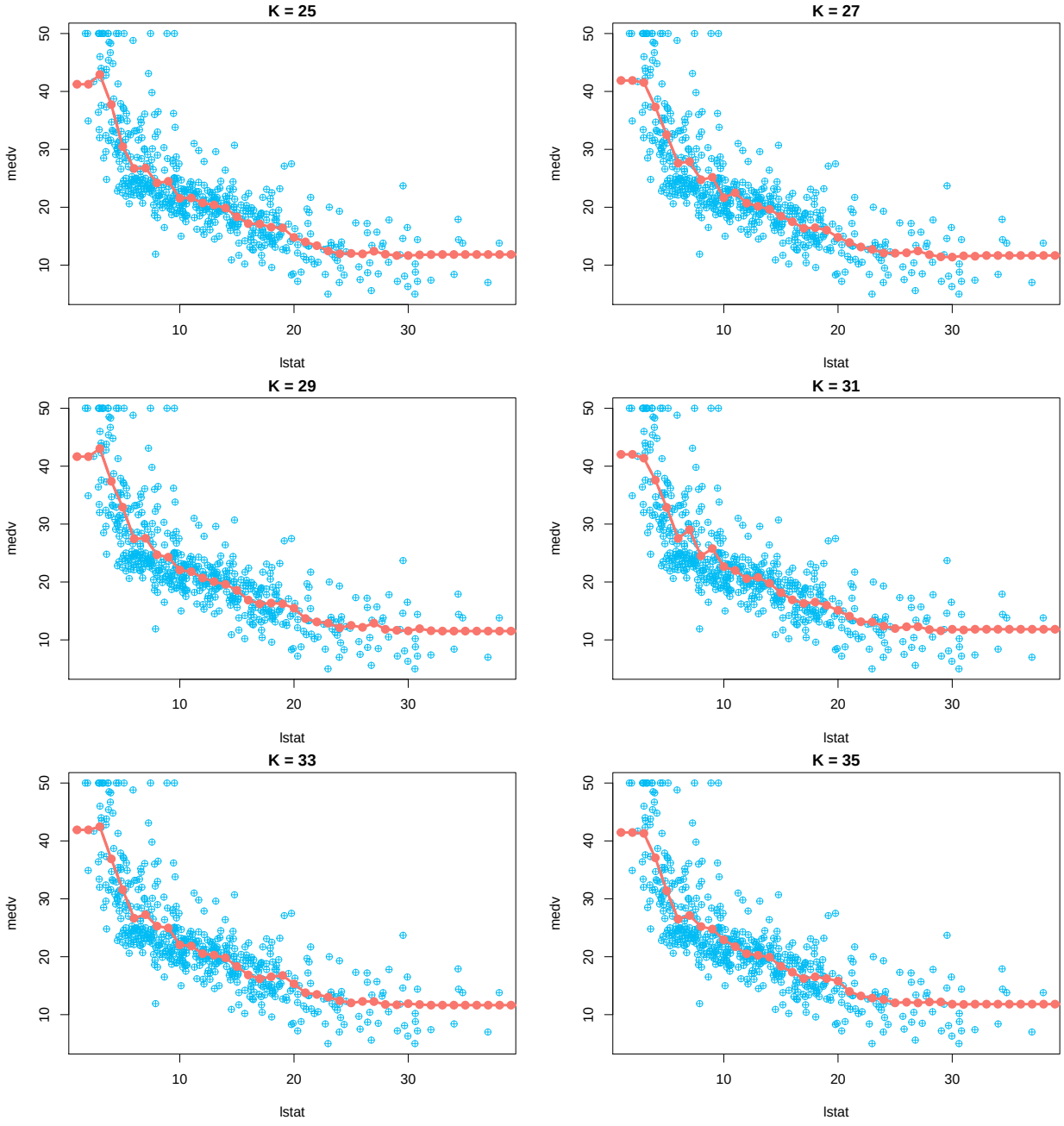


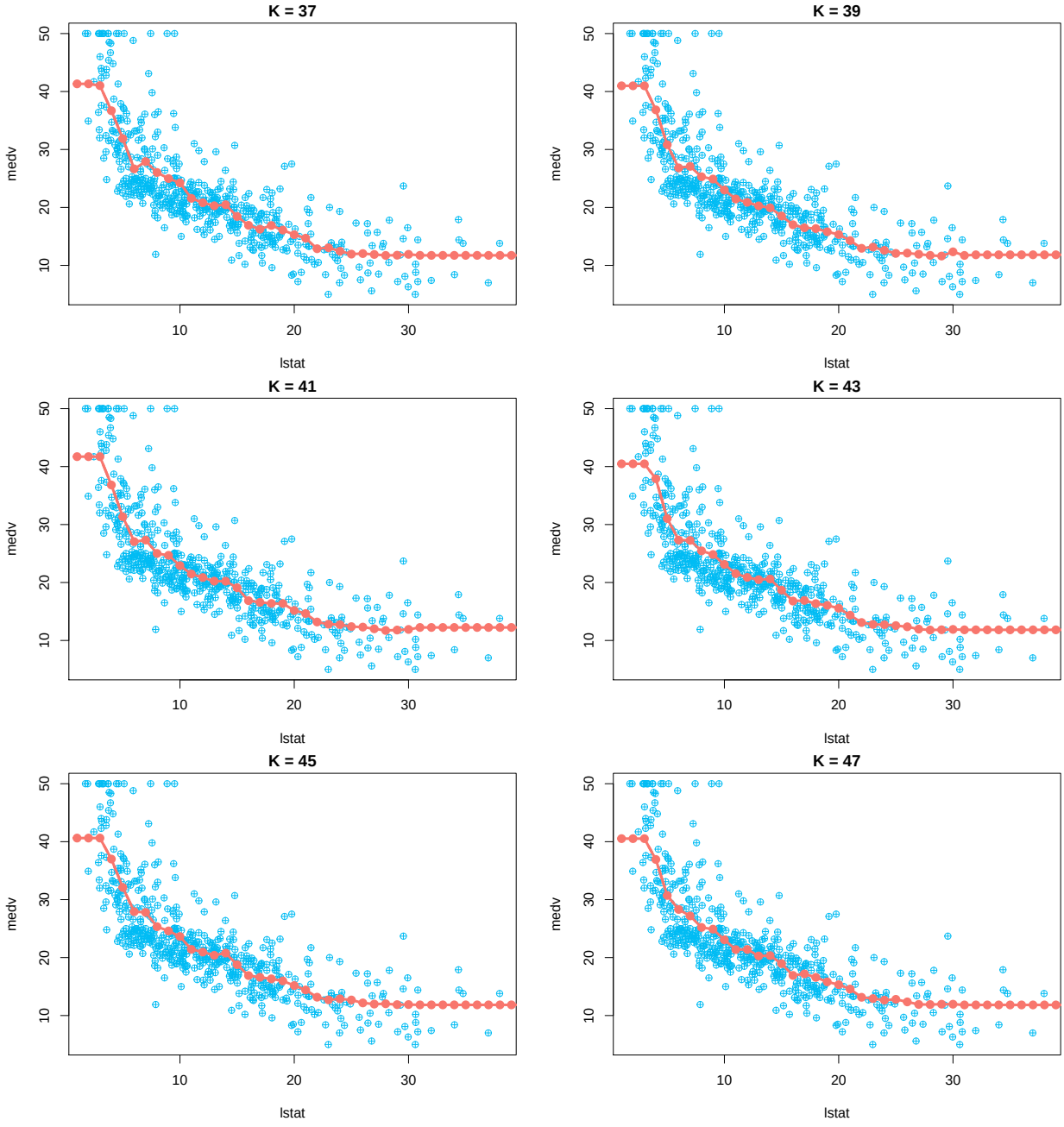
5. Repeat the same exercise using different values of k.

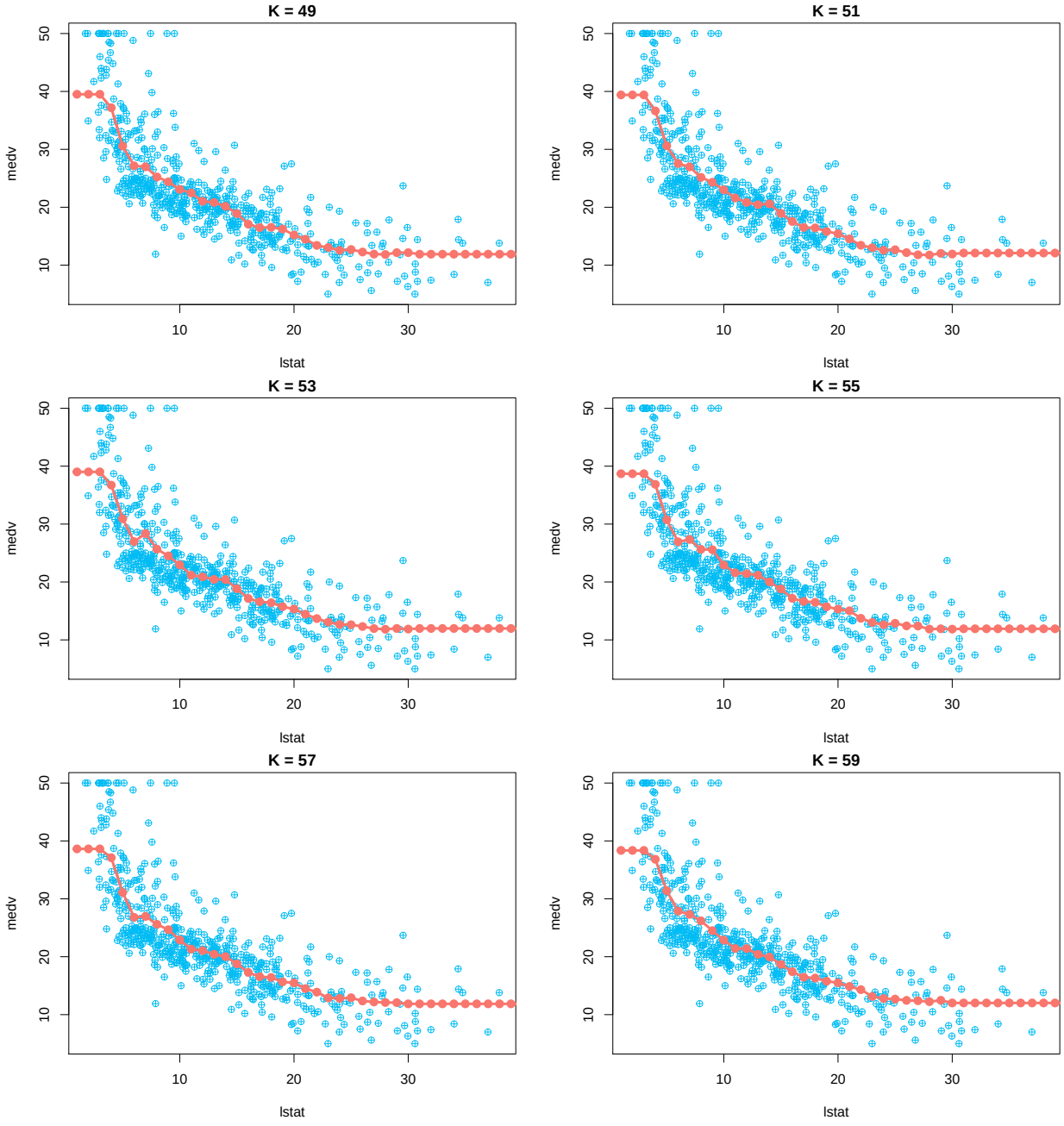
```
for (n in 1:10) {
  a <- seq(from = 1, to = 55, by = 6)
  b <- seq(from = 6, to = 60, by = 6)
  k <- seq(a[n],b[n], by = 2)
  t <- 1:40
  hat_p.2 <- matrix(0,1, ncol = length(k), nrow = length(t))

  for (i in 1:length(t)) {
    for (j in 1:length(k)) {
      hat_p.2[i,j] <- knn.reg(t[i], x=x, y=y, k = k[j])
    }
  }
  # Plot all:
  par(mar = c(4, 4, 1.5, 1.5))
    for (i in 1:length(k)) {
      plot(x, y, pch = 10,xlab = "lstat", ylab = "medv", col = "#00BCF4")
      points(t, as.matrix(hat_p.2[,i]), col="#F8766D" , lwd=3 , pch=19 , type= "l")
      points(t, hat_p.2[,i], col="#F8766D" , lwd=3 , pch=19 , type= "b")
      title(paste("K =", k[i]))
    }
}
```


K = 1
K = 3
K = 5
K = 7
K = 9
K = 11
medv
lstat
10
20
30
40
50


K = 13
K = 15
K = 17
K = 19
K = 21
K = 23
medv
lstat
10
20
30
40
50


K = 25
K = 27
K = 29
K = 31
K = 33
K = 35
medv
lstat
10
20
30
40
50


K = 37
K = 39
K = 41
K = 43
K = 45
K = 47
medv
lstat
10
20
30
40
50


K = 49
K = 51
K = 53
K = 55
K = 57
K = 59
medv
lstat
10
20
30
40
50