Nonparametric Regression

Crowding on Mt. Everest

Source: Nirma Purjal Project Possible Ltd.

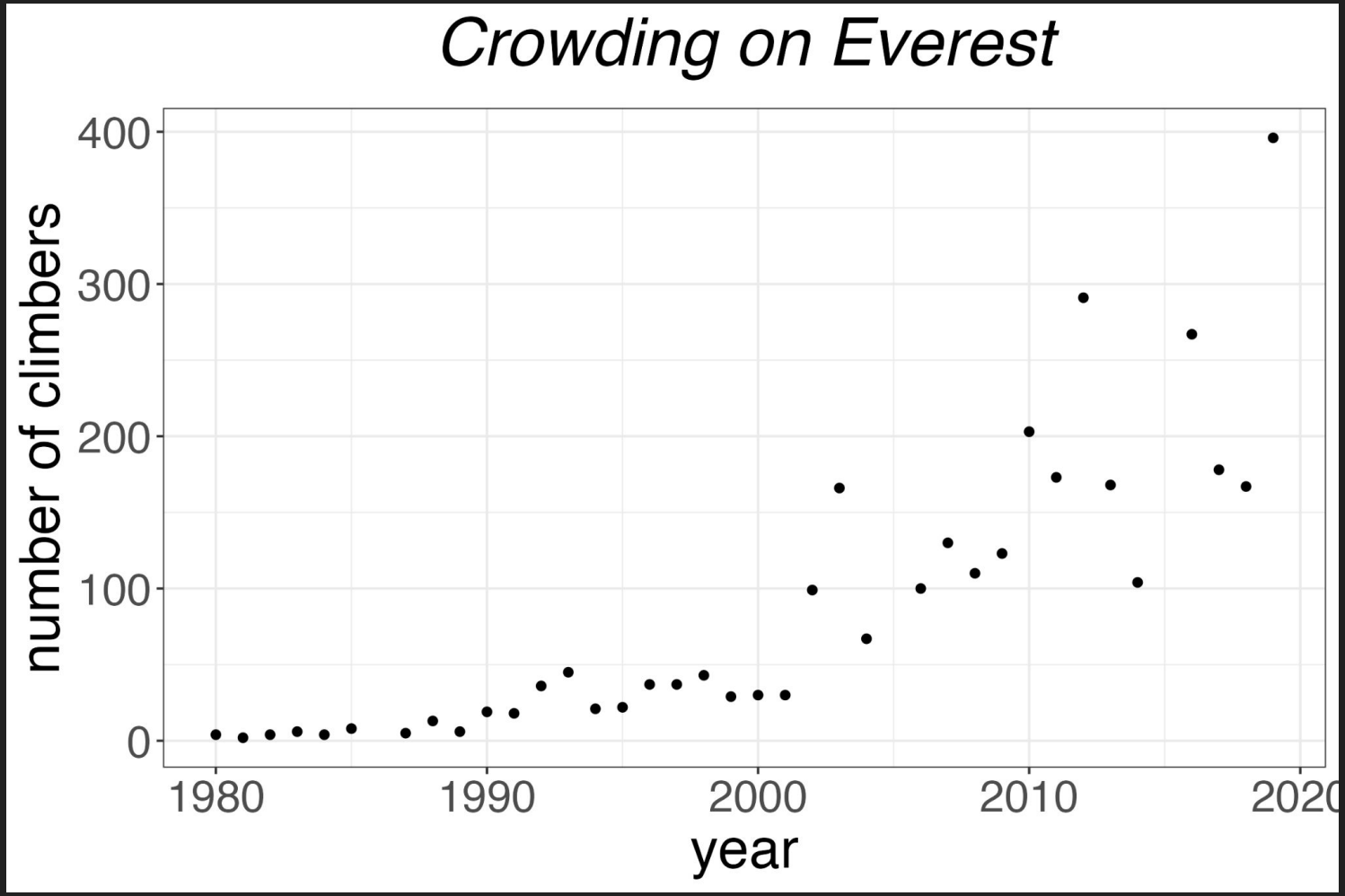
Mt. Everest in the news

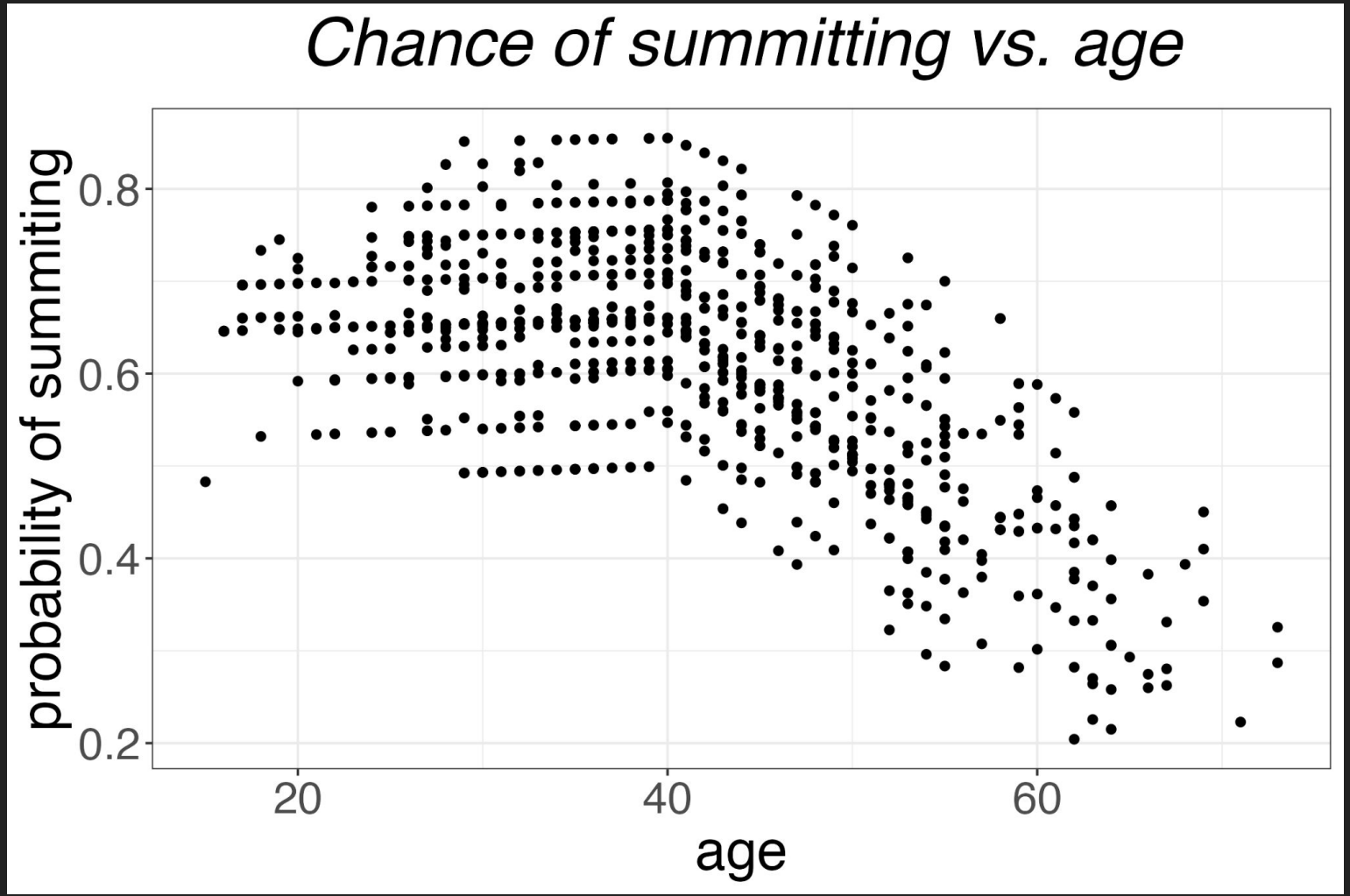
Source: GQ

Policy Q: Should the Nepali government limit Everest permits?

- Safety concerns for climbers vs. economic benefits for Nepal

- How bad has the crowding gotten?


Crowding on Everest
number of climbers
0
100
200
300
400
1980
1990
2000
2010
2020
year


Chance of summitting vs. age
probability of summiting
0.8
0.6
0.4
0.2
20
40
60
age

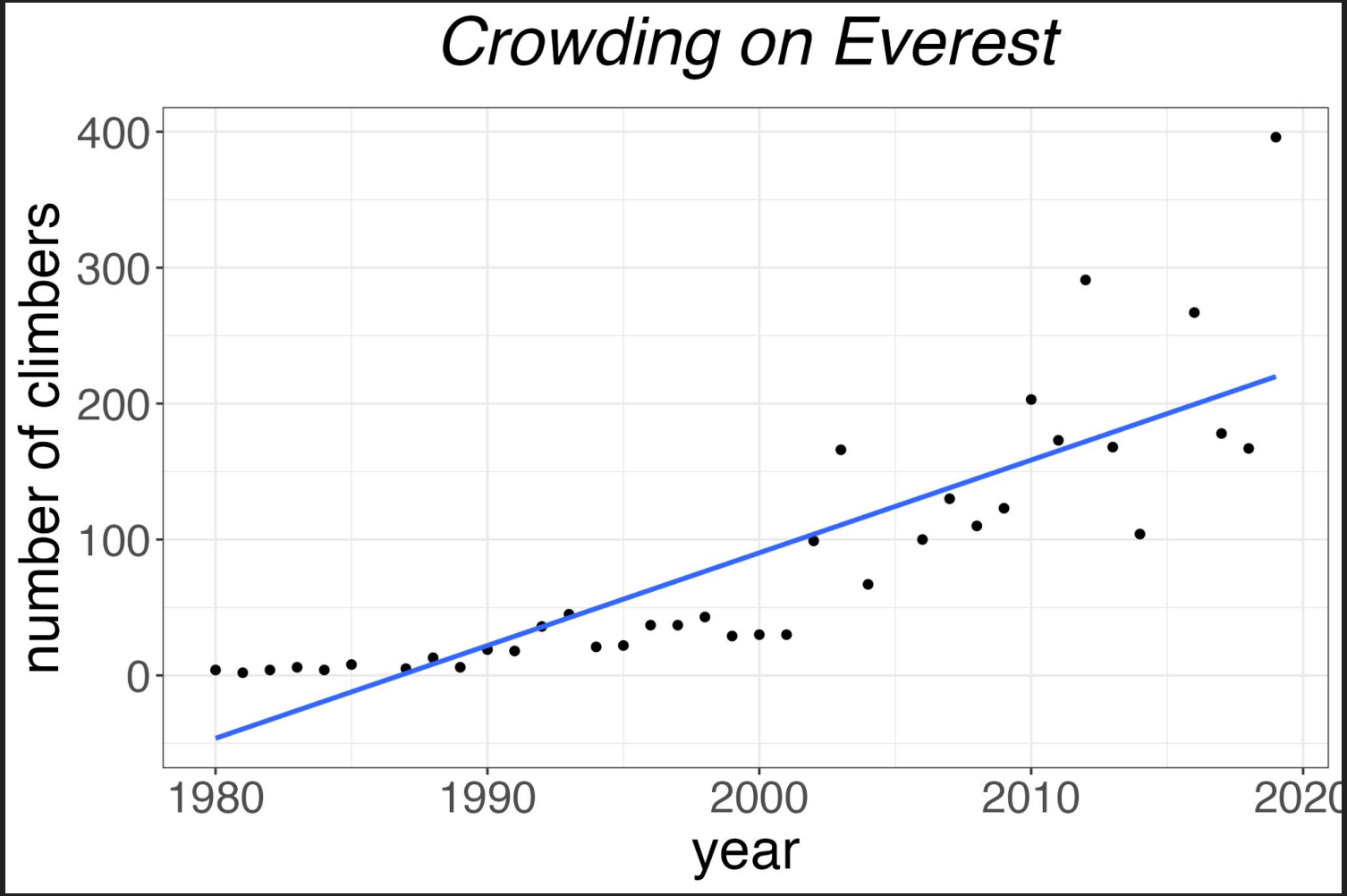
The regression problem

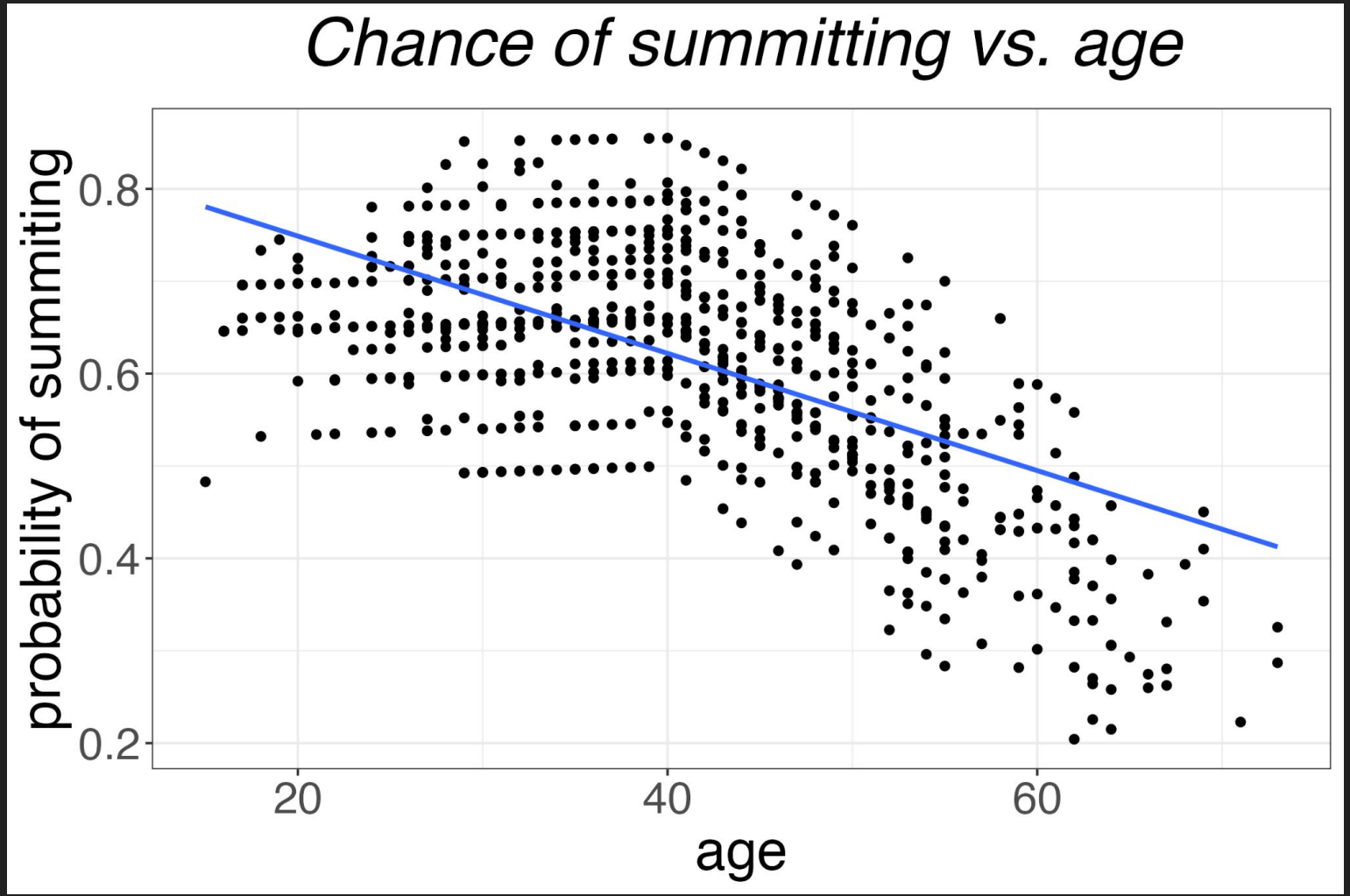
$$y = g(x) + e$$

$$E(e) = 0$$

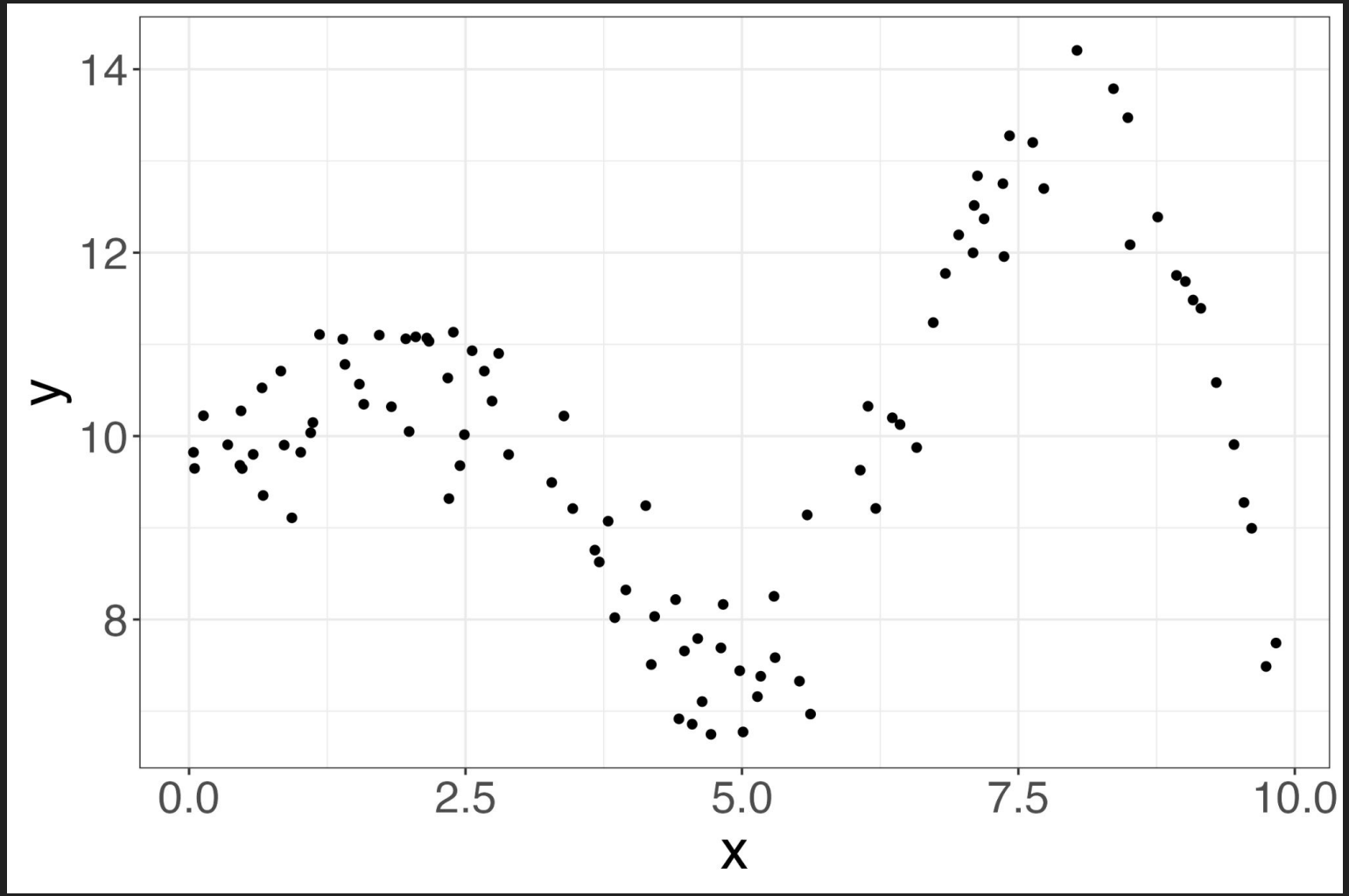
$$y = g(x) + e$$

$$g(x) = \beta_0 + \beta_1 x$$


Crowding on Everest
number of climbers
400
300
200
100
0
1980
1990
2000
2010
2020
year


Chance of summitting vs. age
probability of summiting
0.8
0.6
0.4
0.2
20
40
60
age

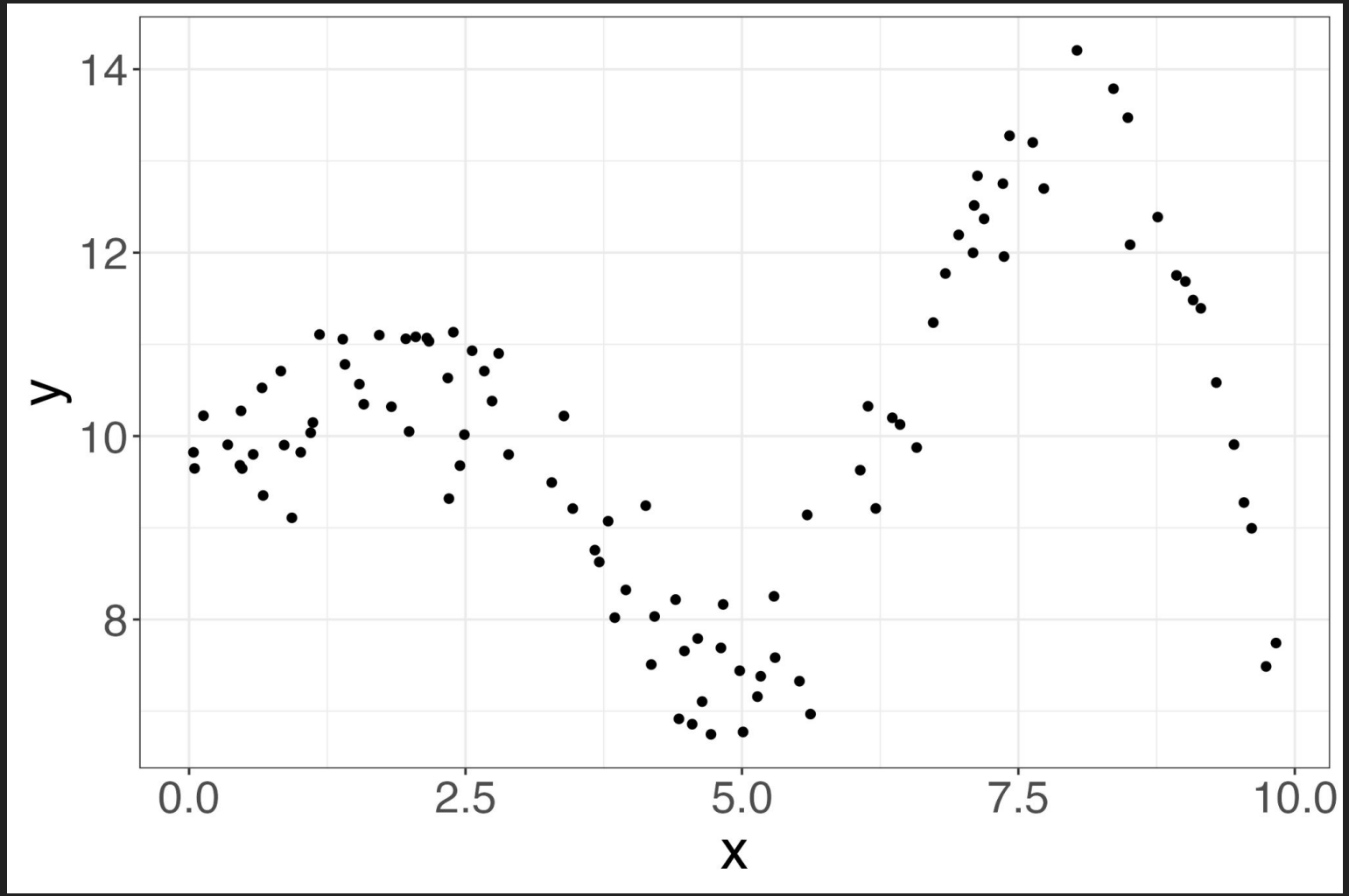
Need a flexible alternative to OLS


14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

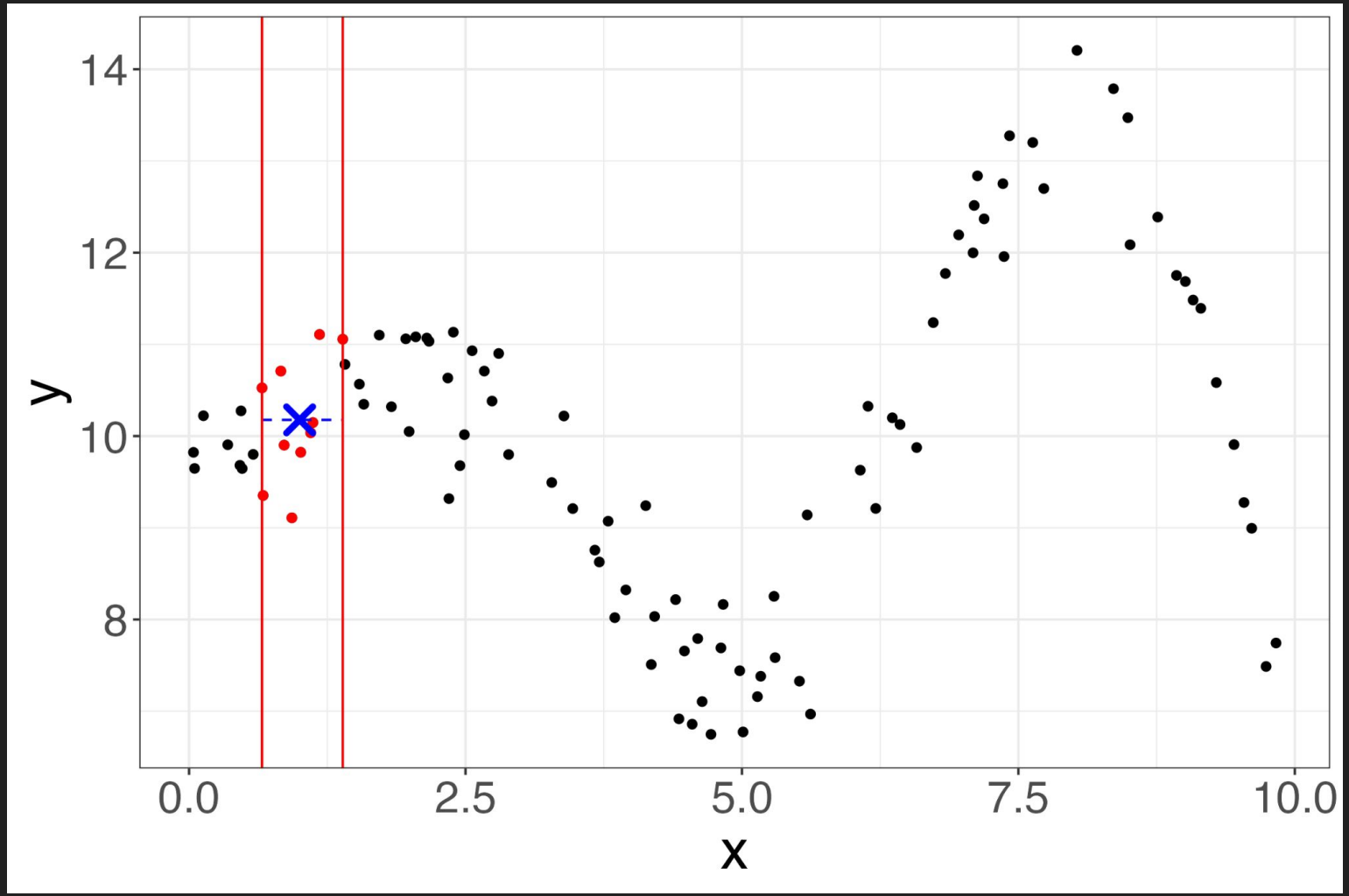
Could do polynomial regression...

Ex: $g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_{p-1} x^{p-1}$

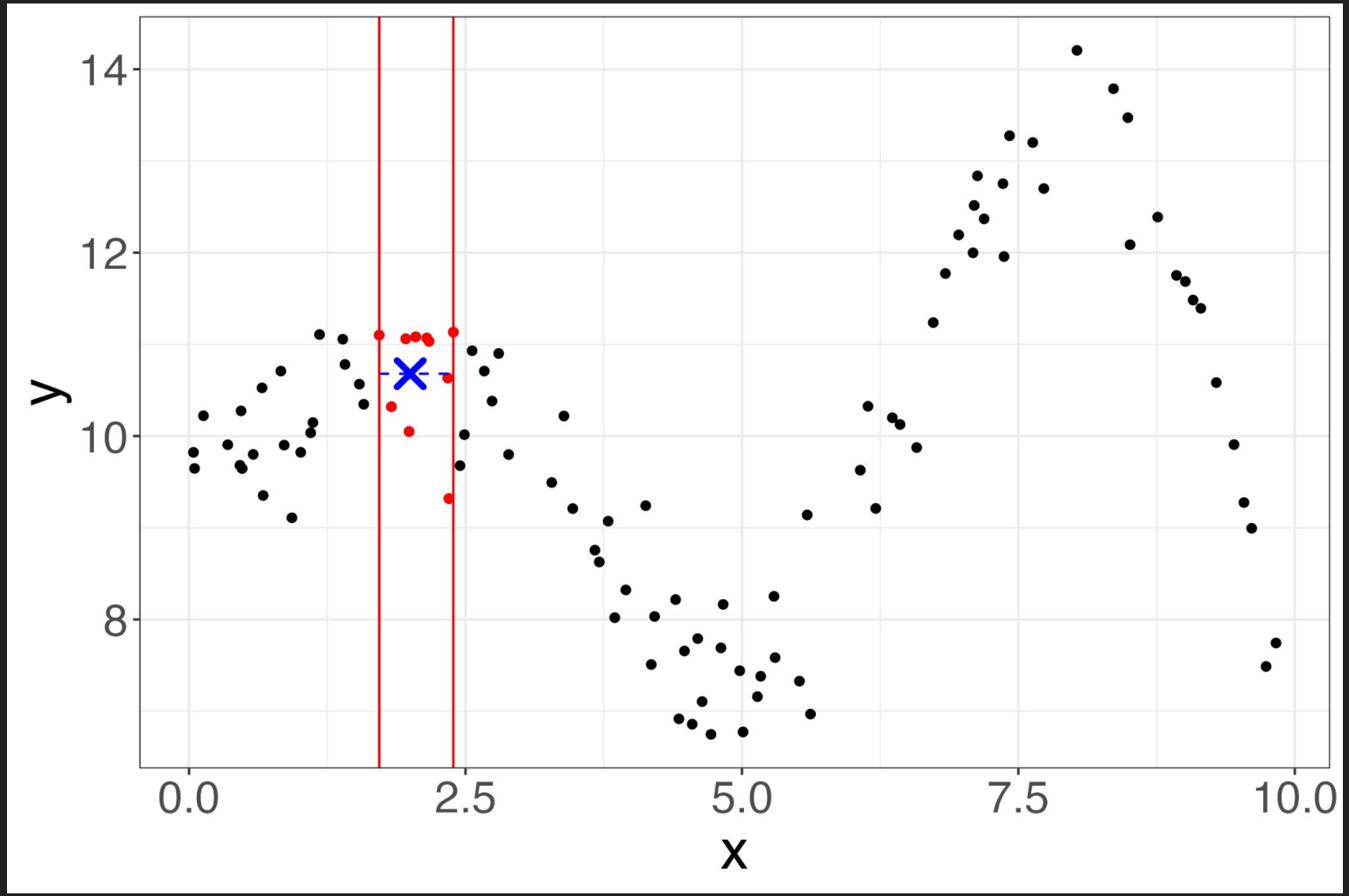
But let’s try something simpler


14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

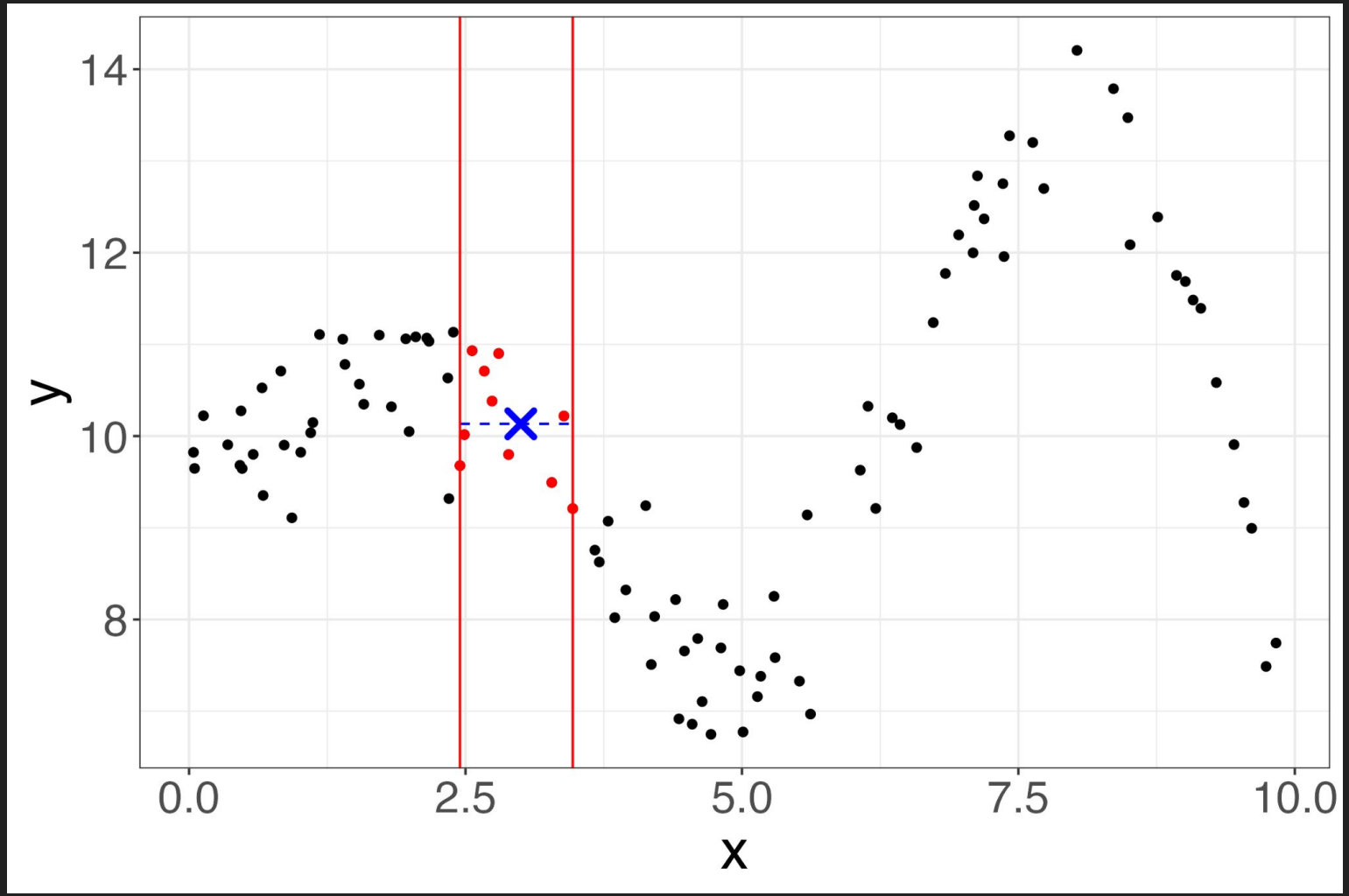
Prediction when x=1

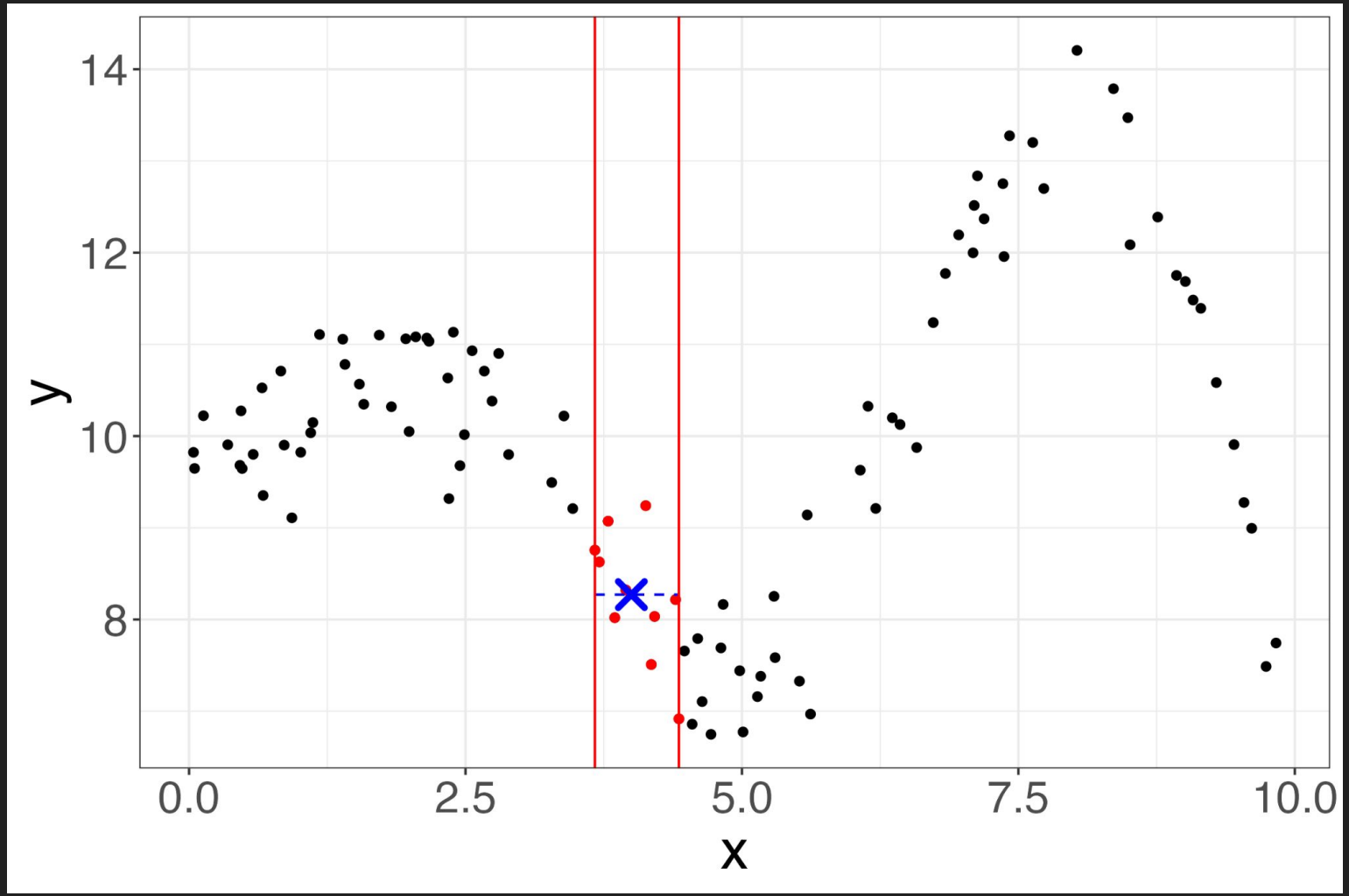

14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

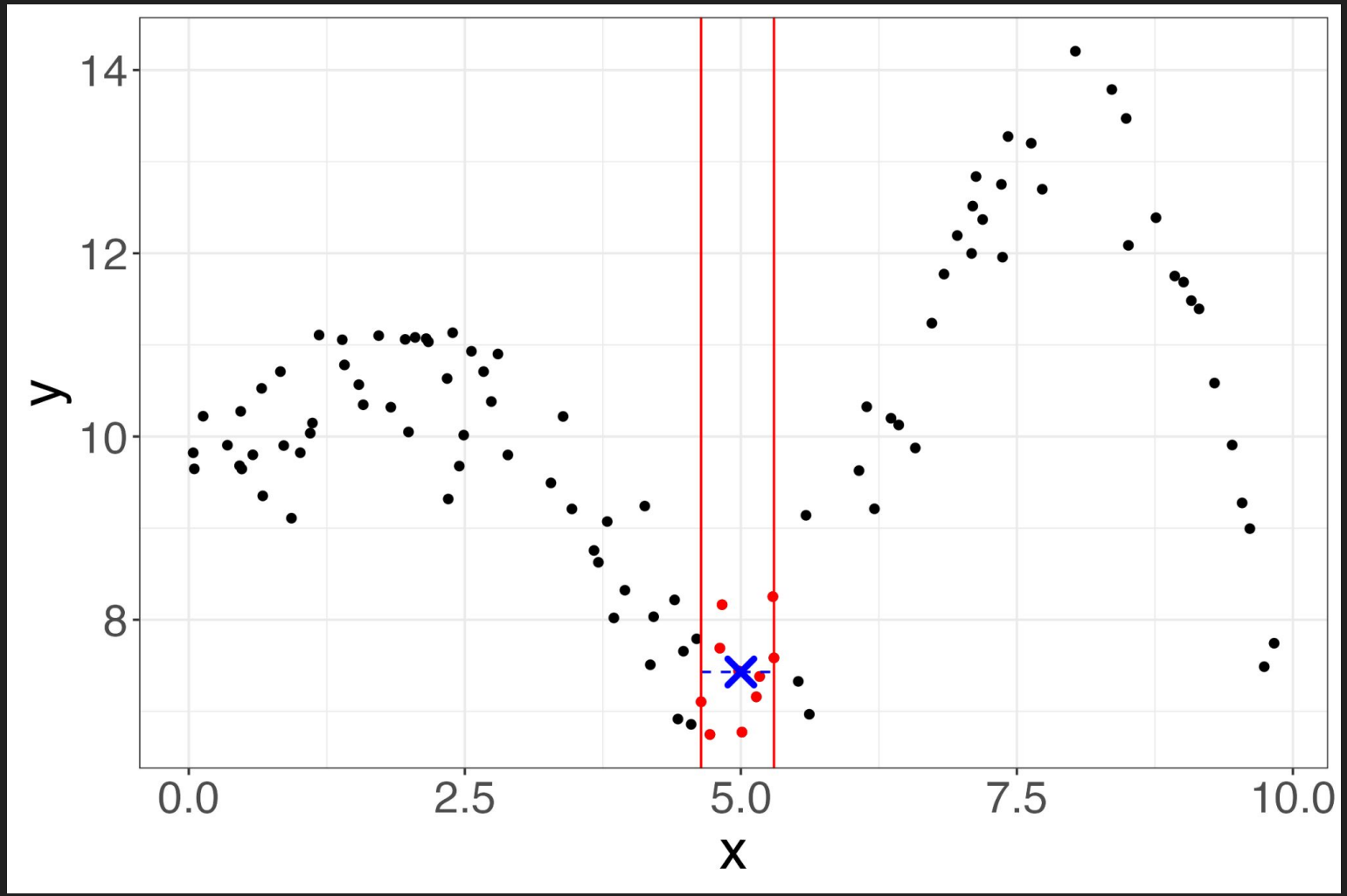
Prediction when x=2

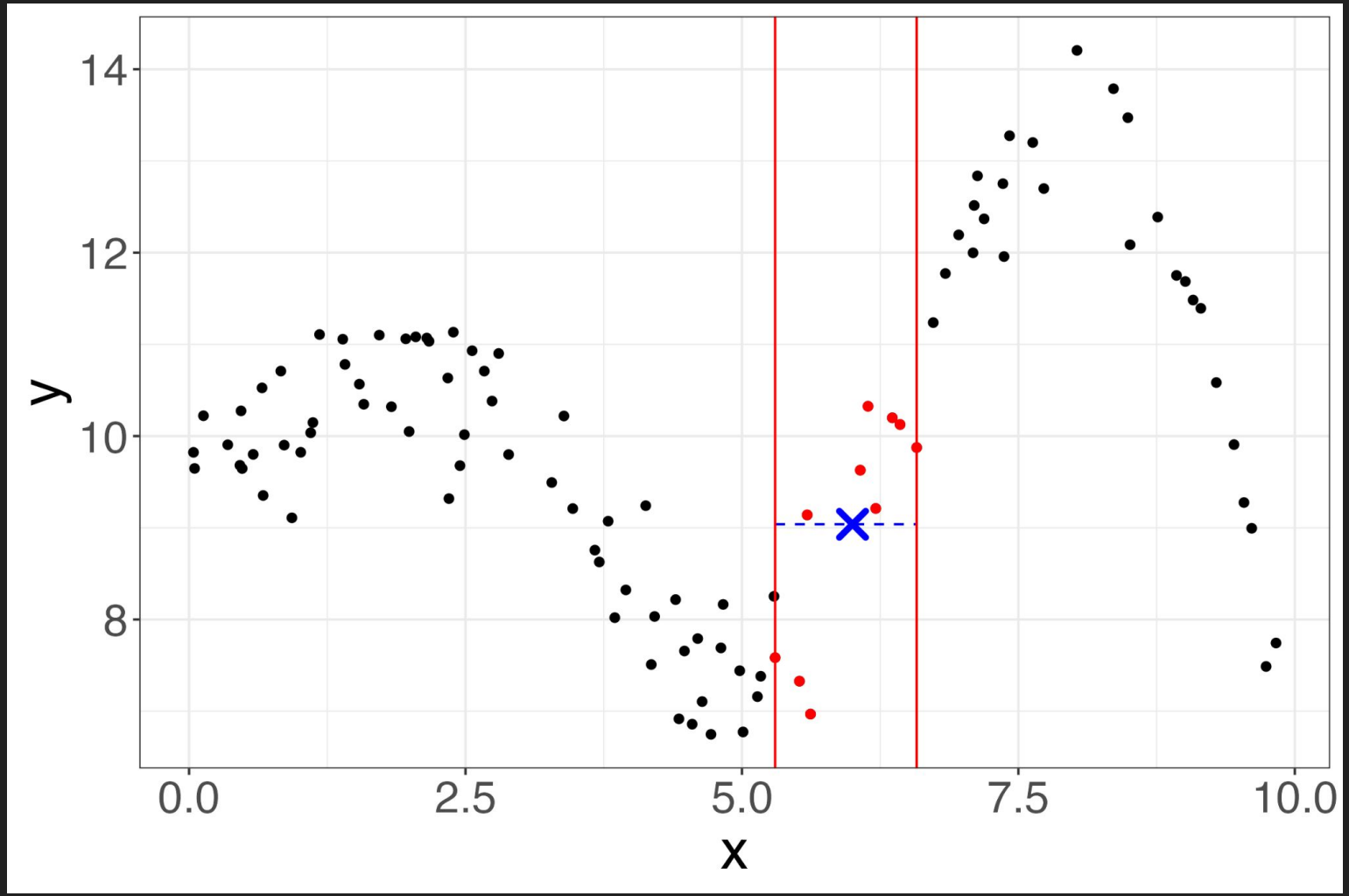

14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

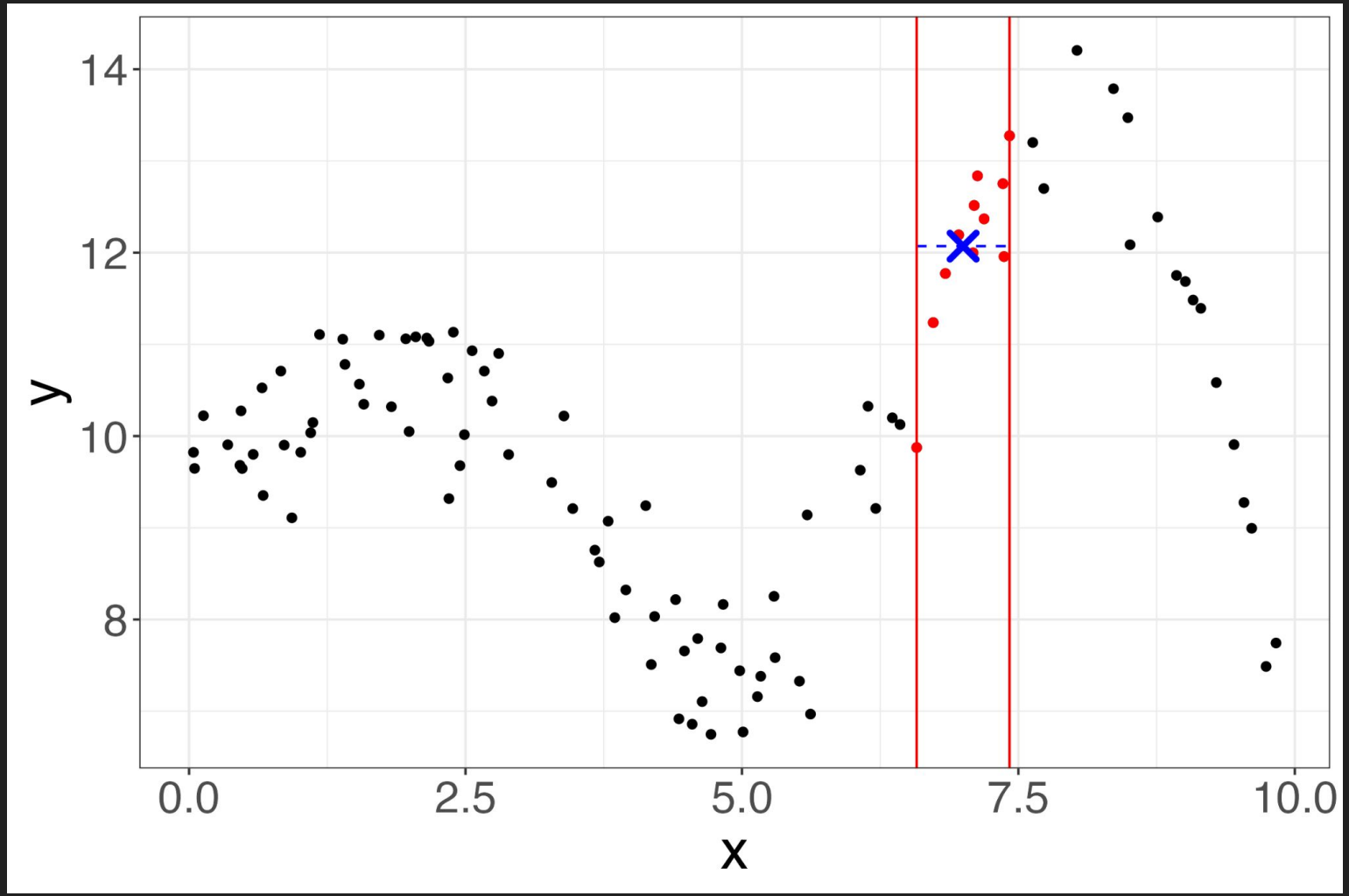
Keep going...

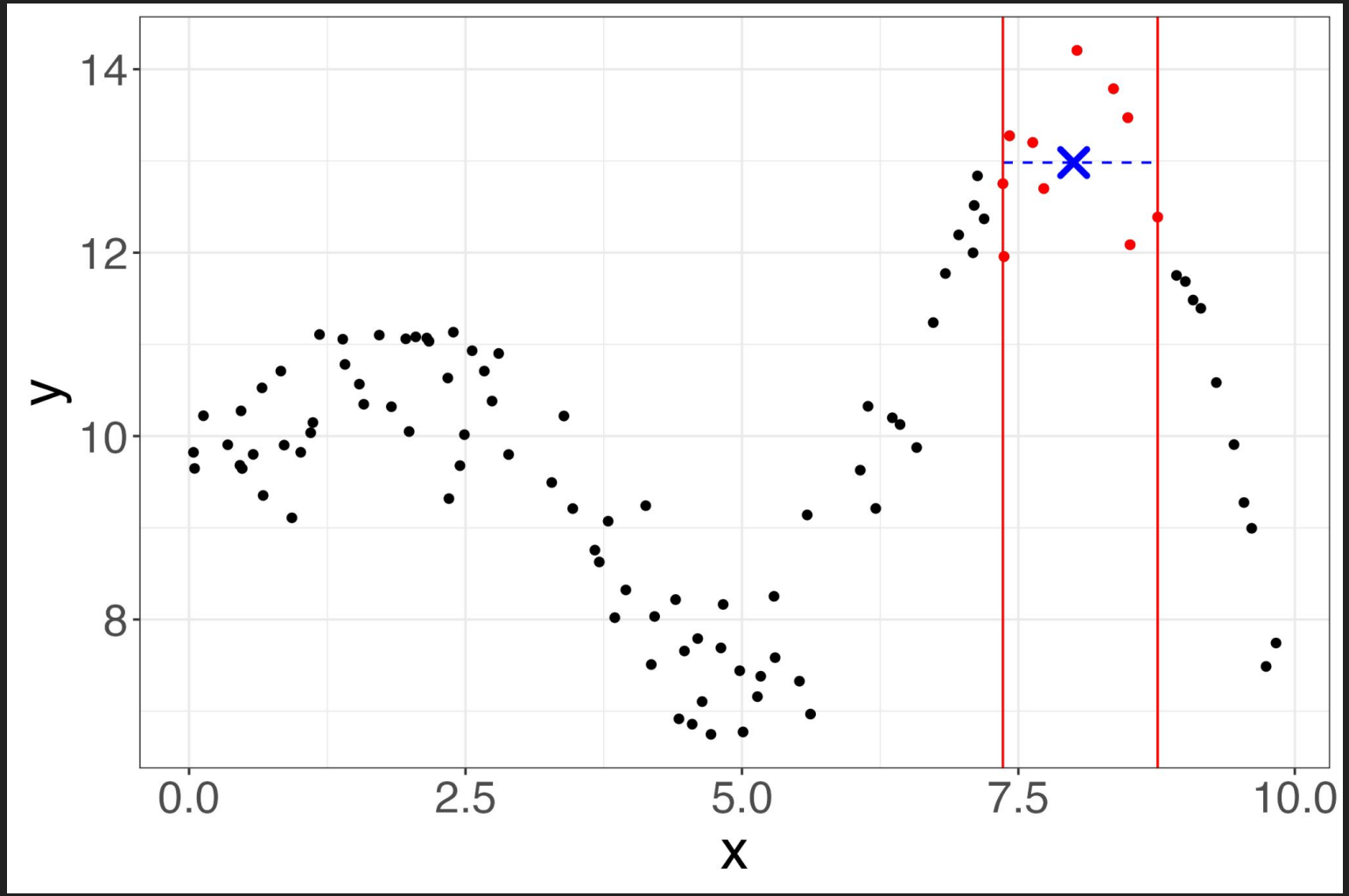

14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

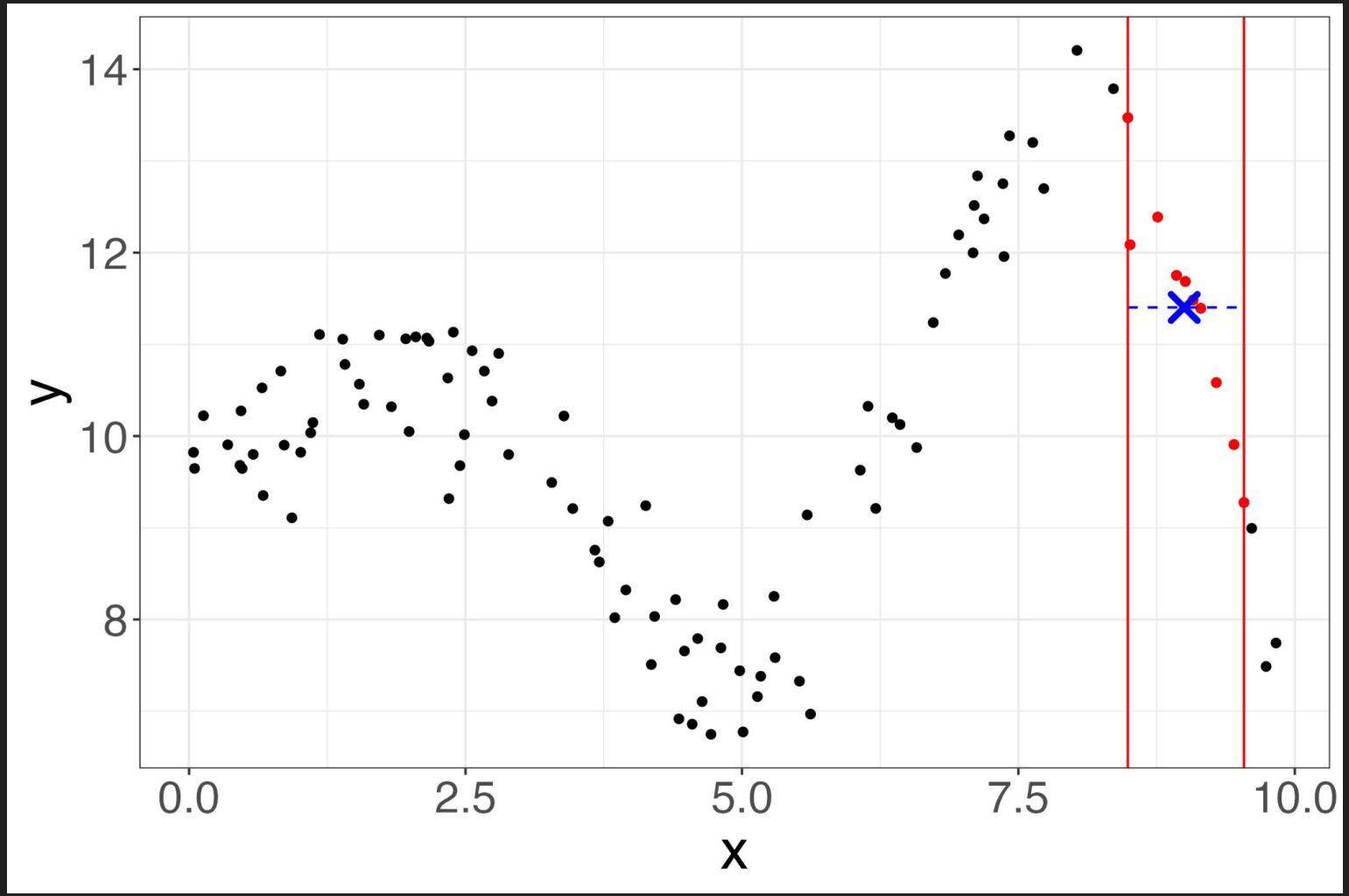

14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

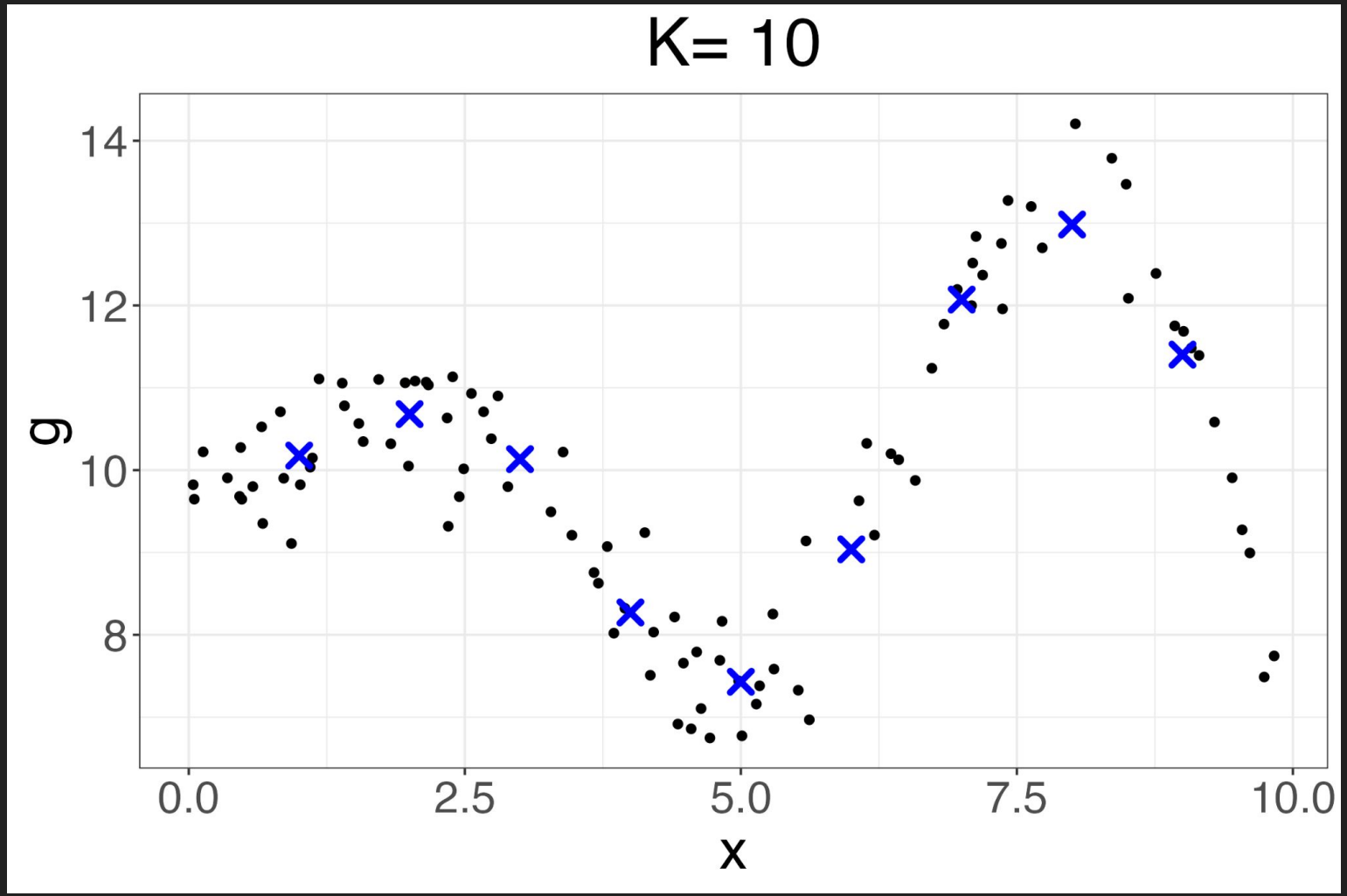

14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

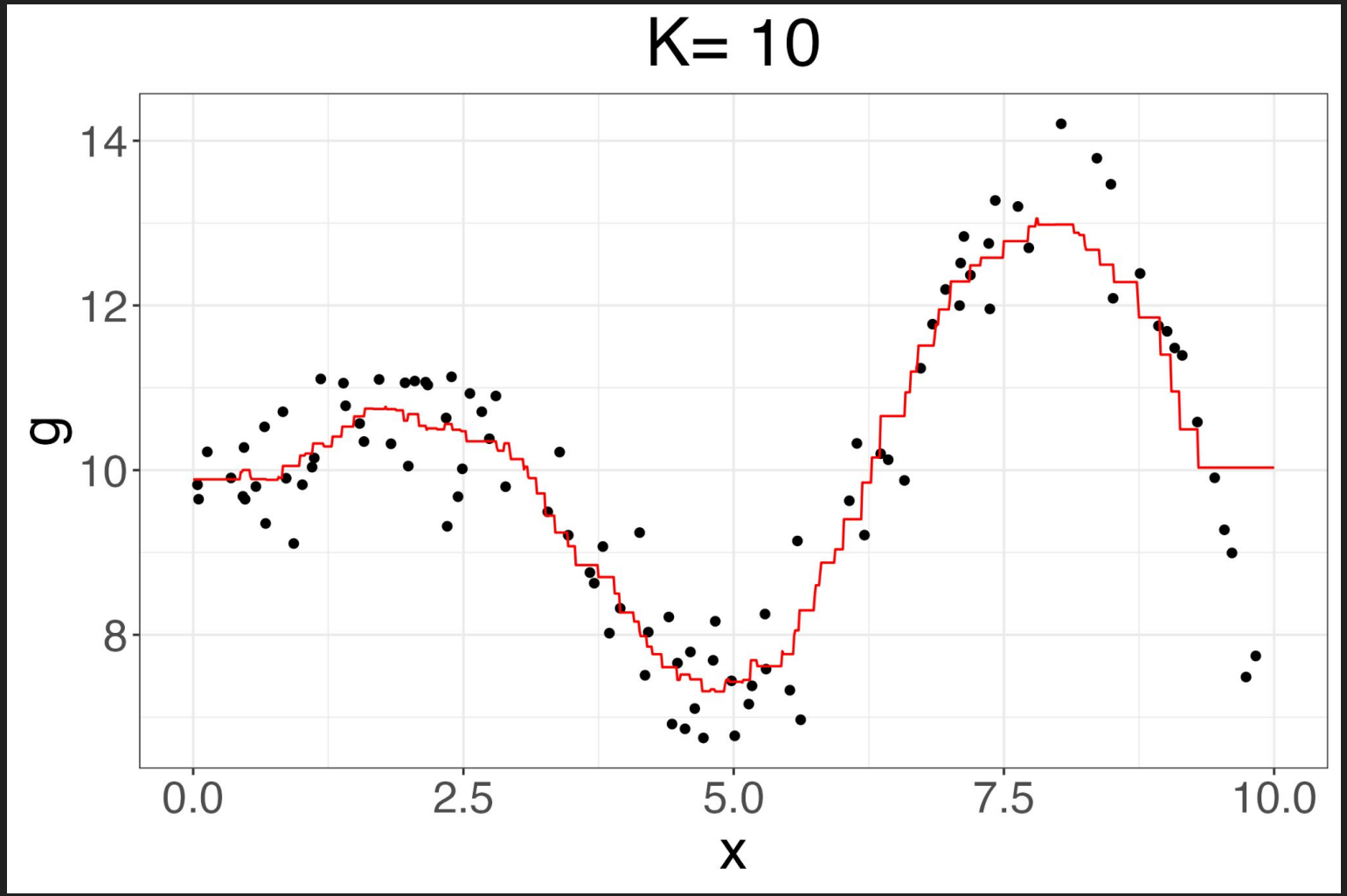

14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x


14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x


14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

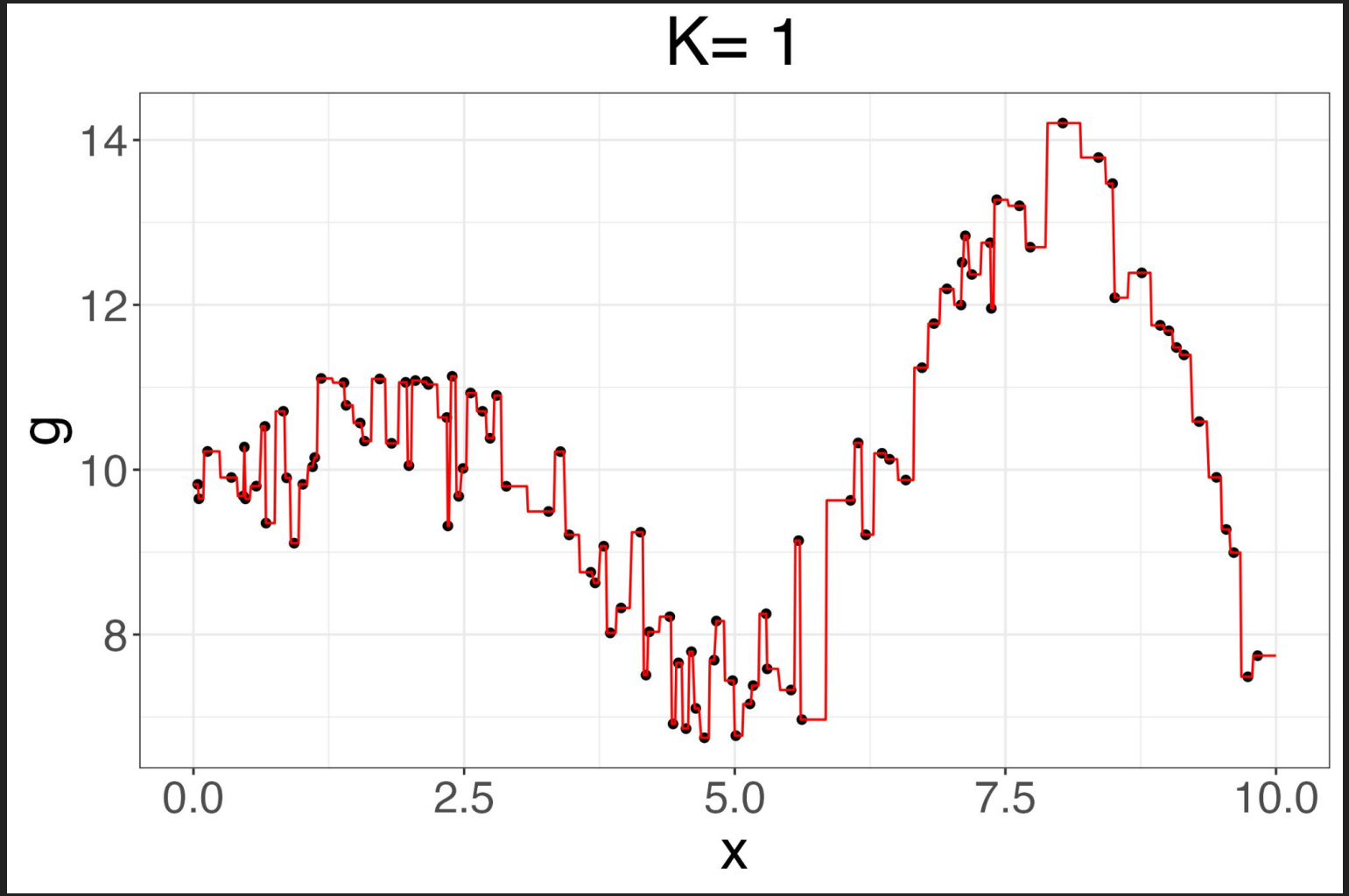

14
12
10
8
y
0.0
2.5
5.0
7.5
10.0
x

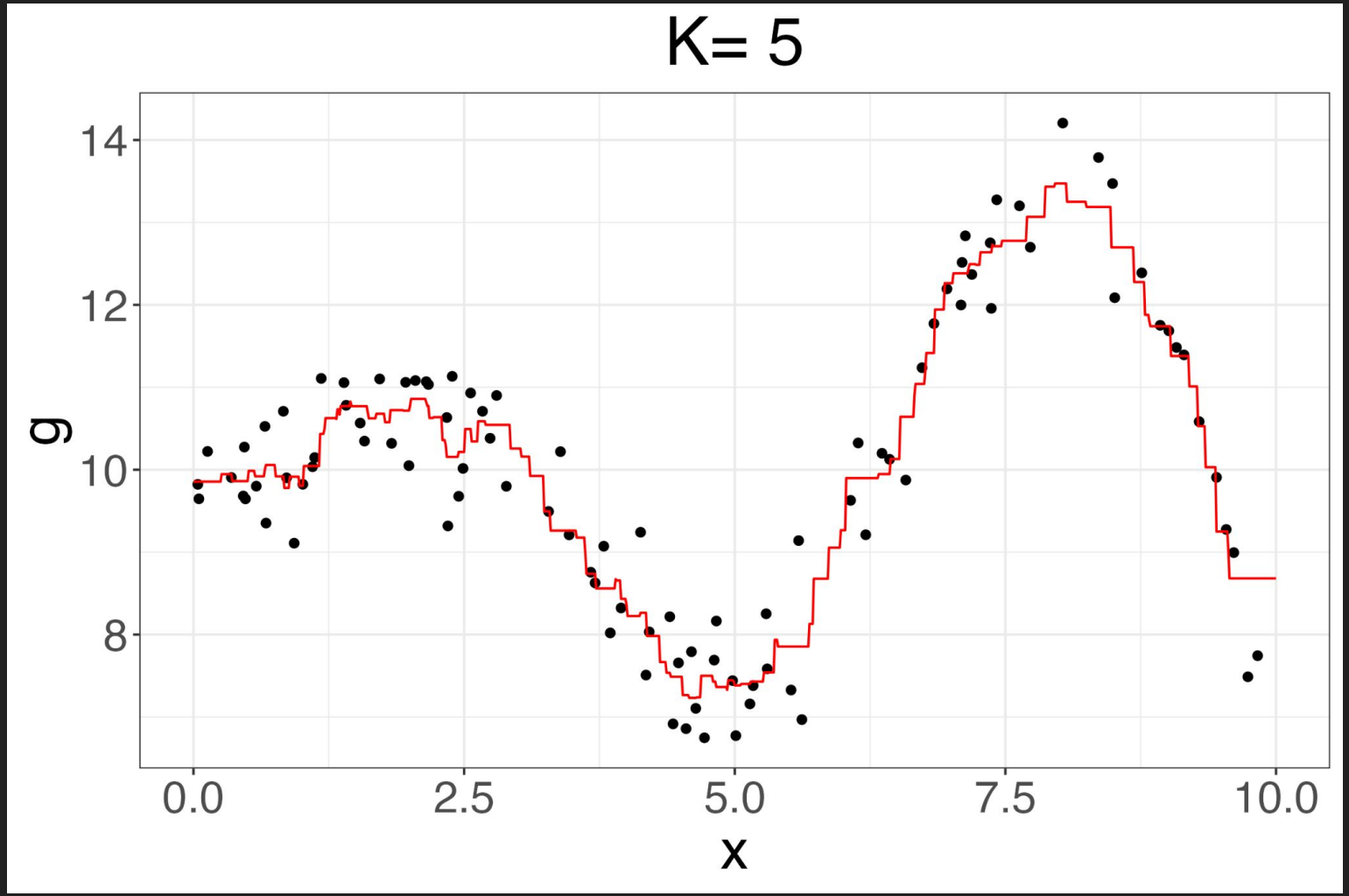

K= 10
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x

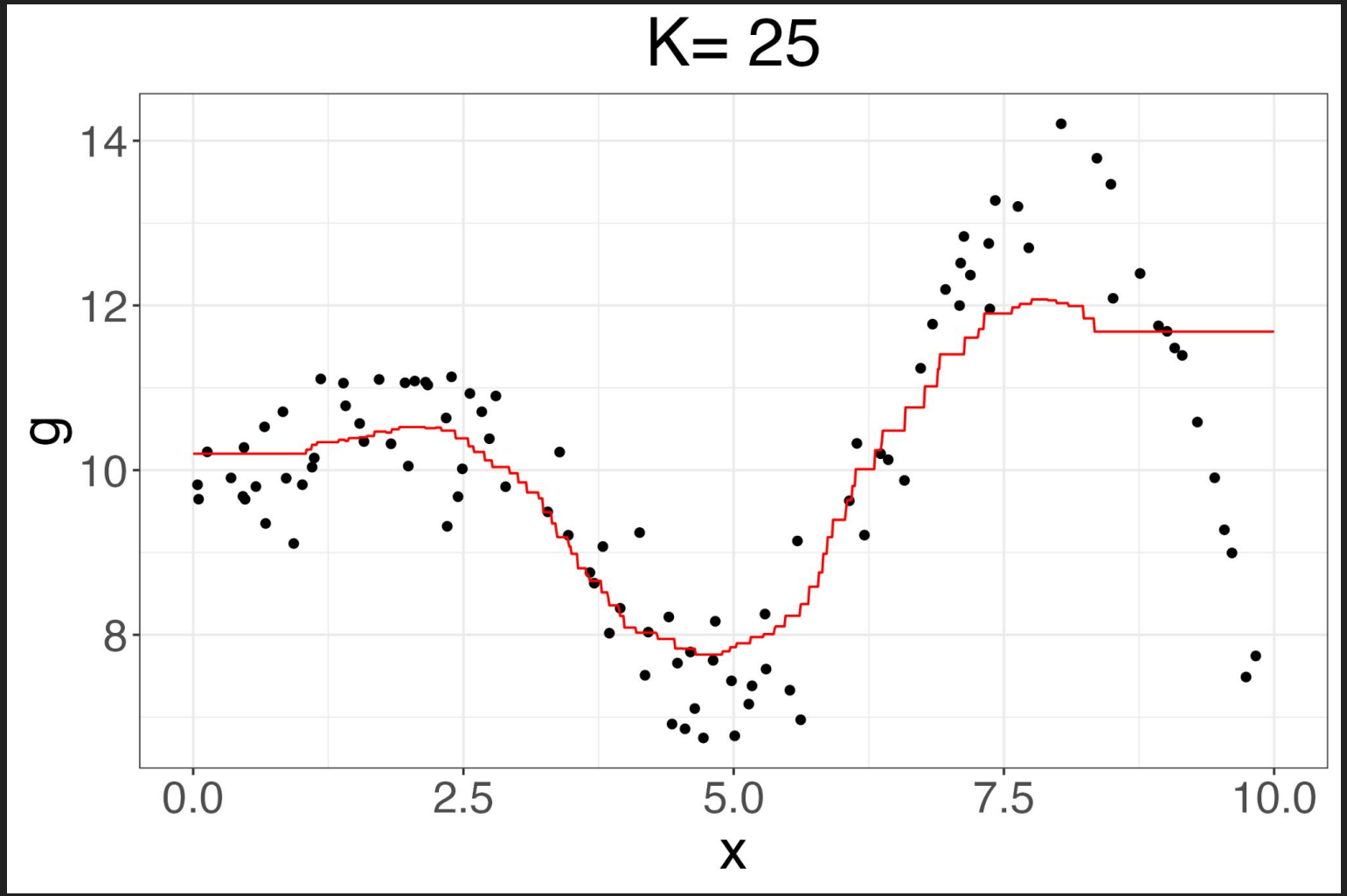

K= 10
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x

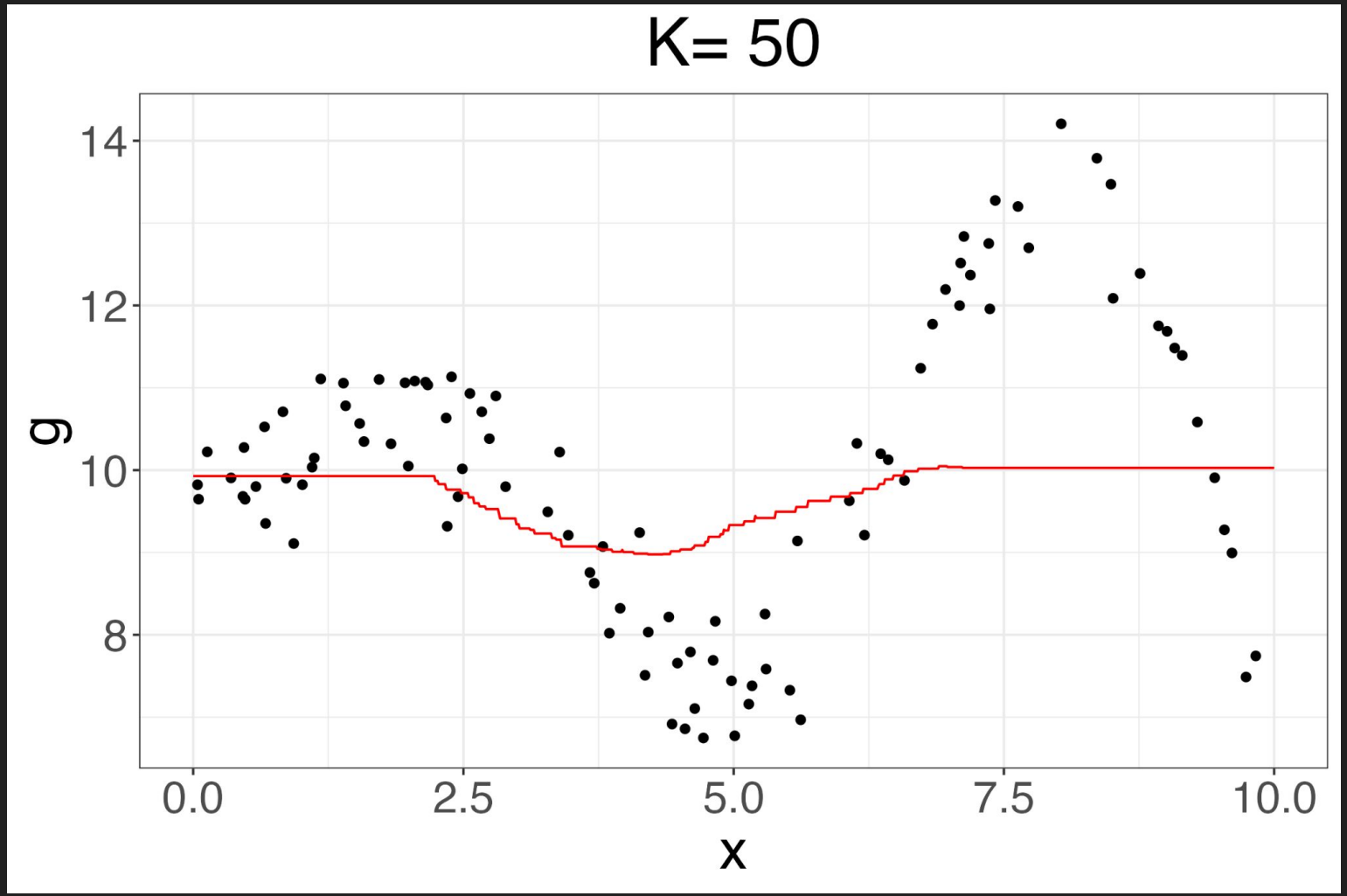
K-nearest neighbor regression (Fix and Hodges 1951)

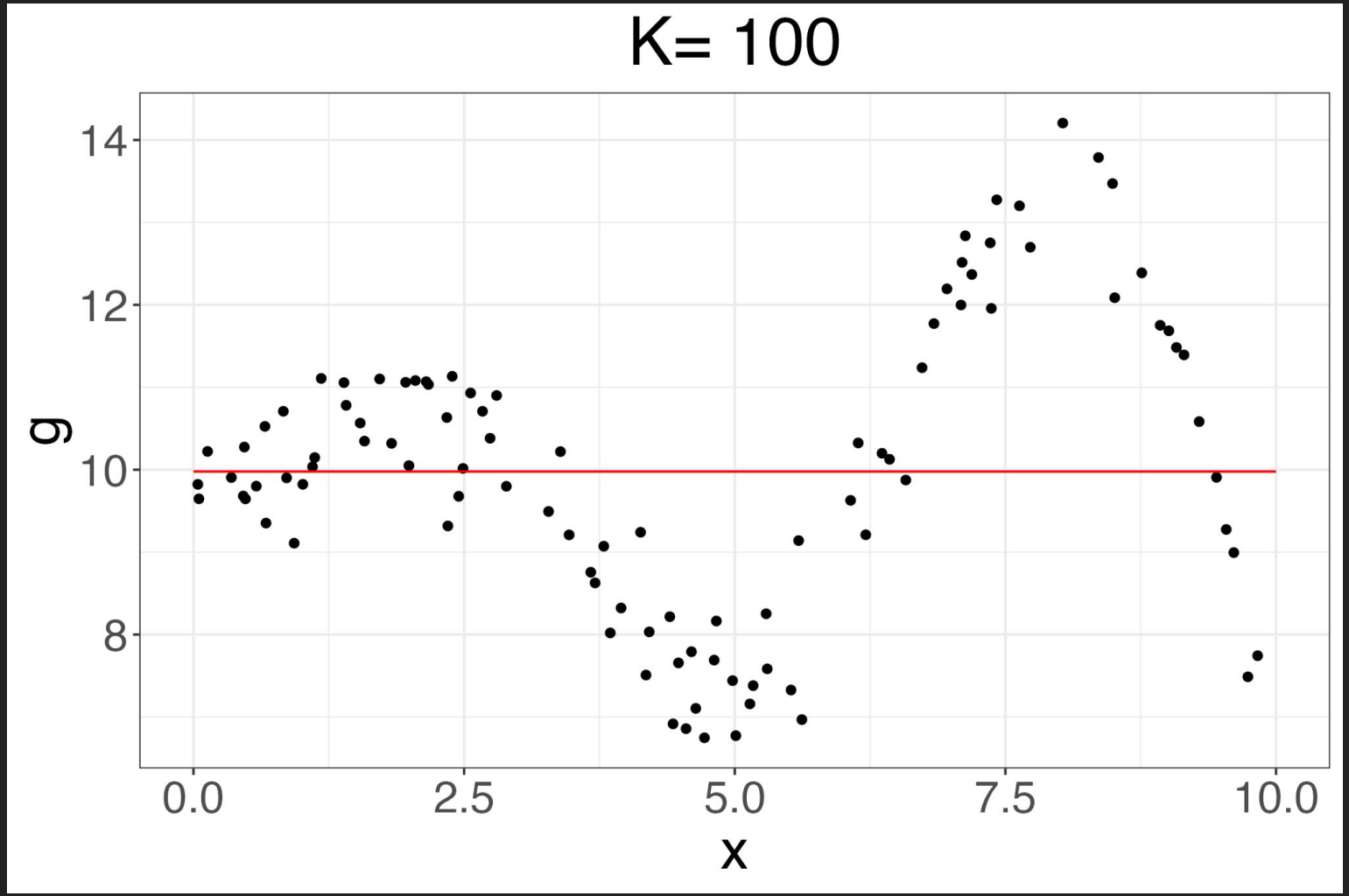
Why 10 neighbors?


K= 1
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x


K= 5
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x


K= 25
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x


K= 50
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x


K= 100
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x

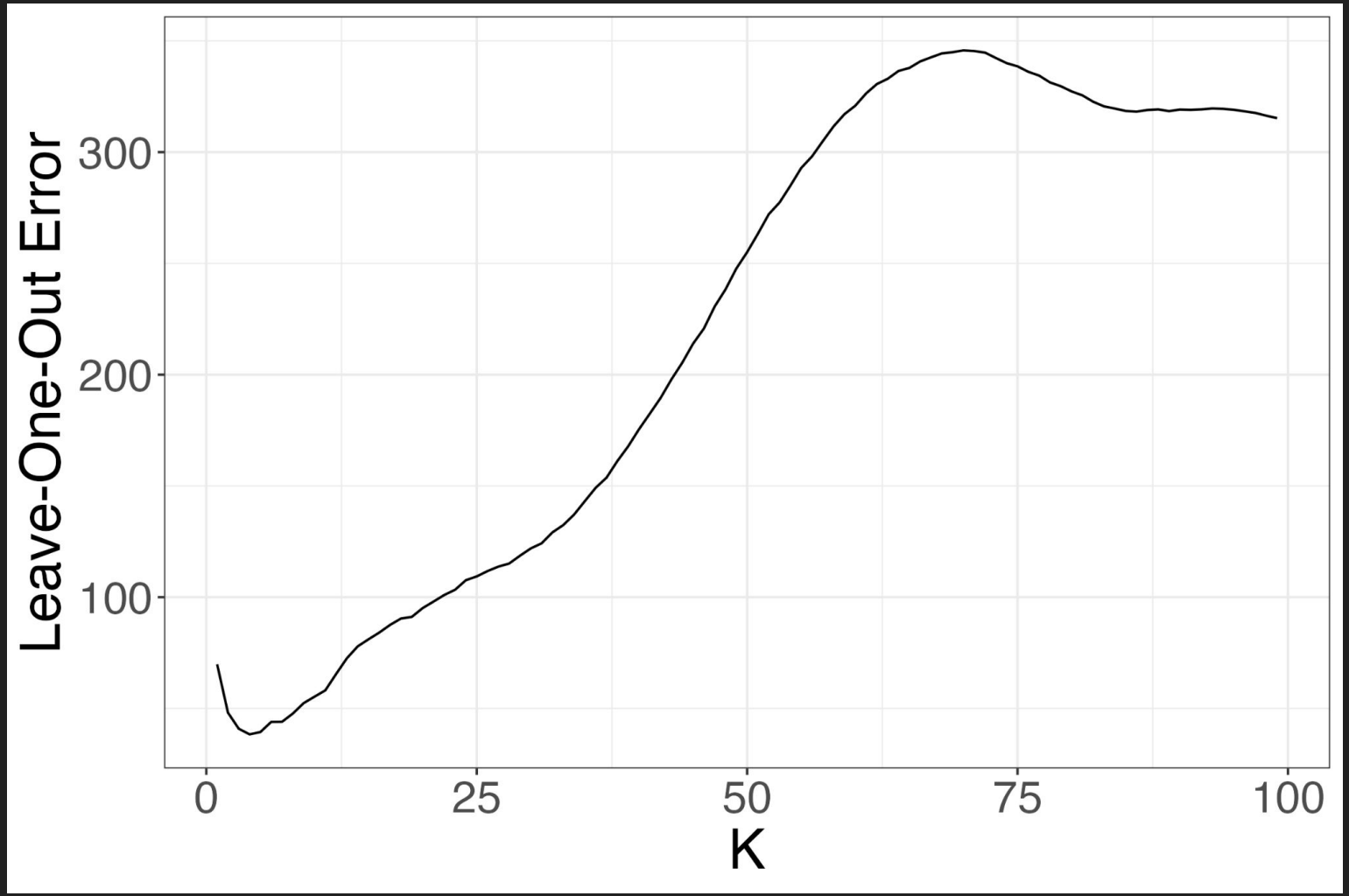
Overfitting
vs.
Oversmoothing

Bias-Variance Tradeoff

Choosing the right K

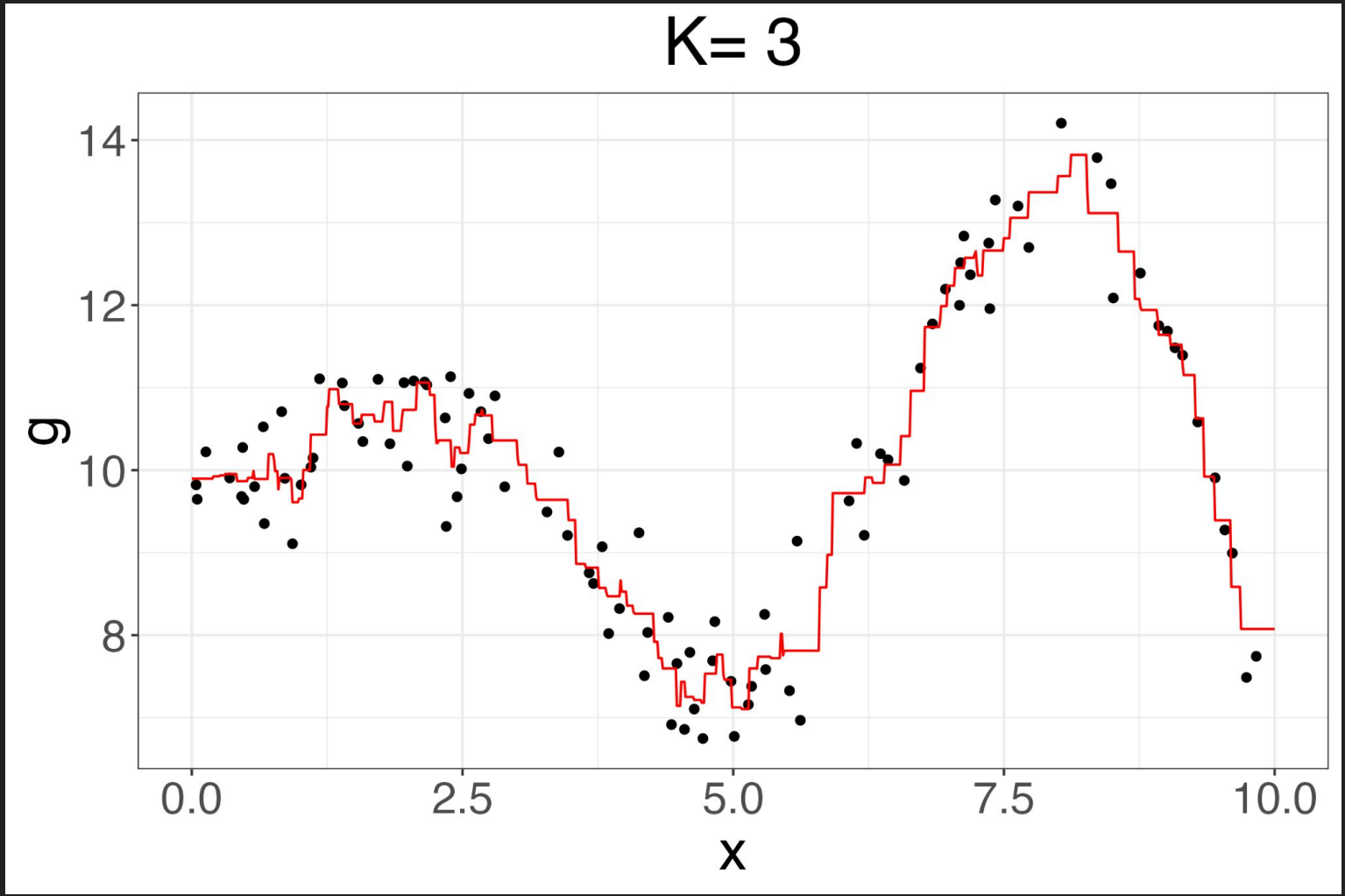
- Visual inspection

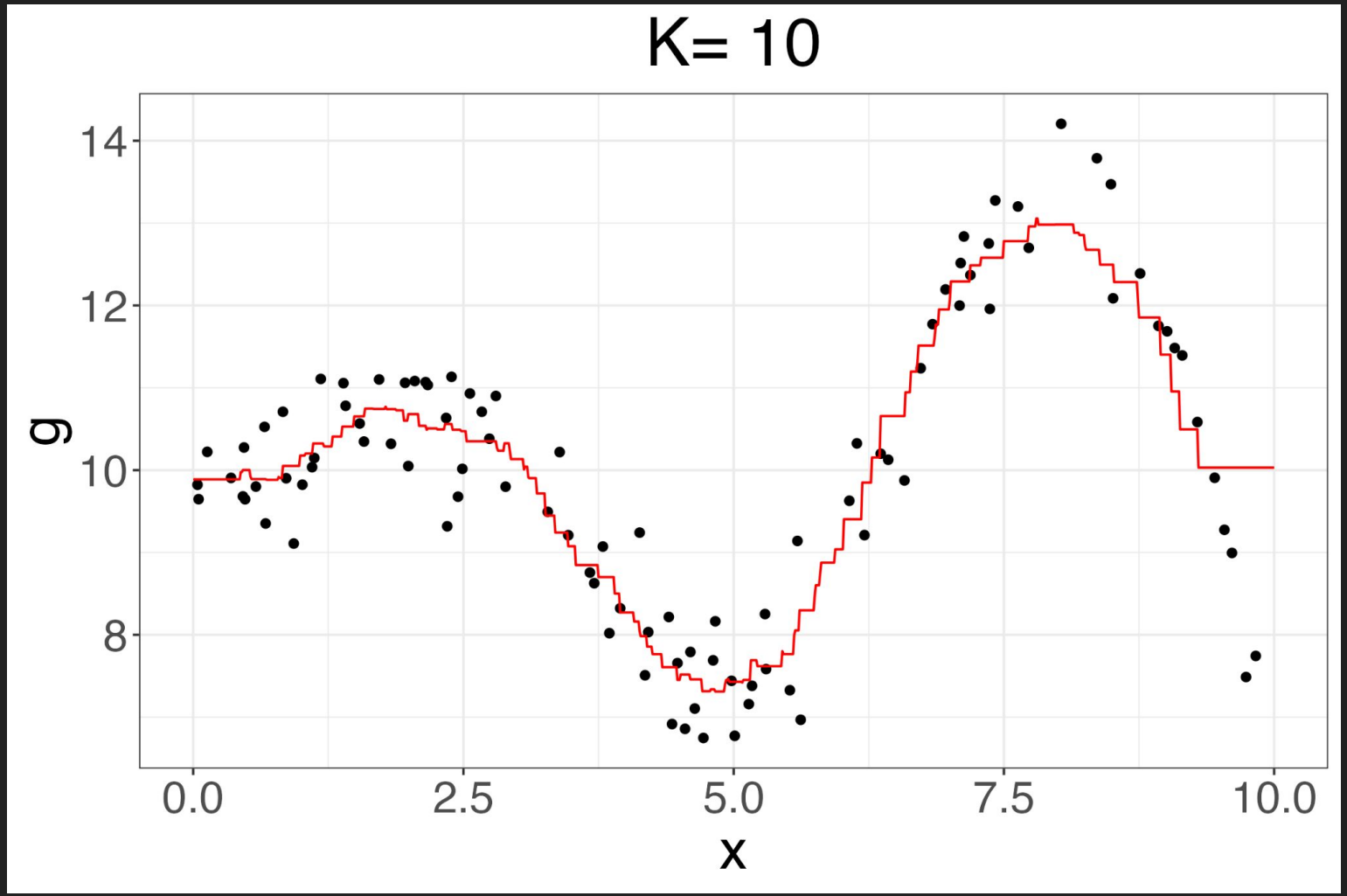
- Cross-validation: minimizing error on a validation set


Leave-One-Out Error
300
200
100
0
25
50
75
100
K

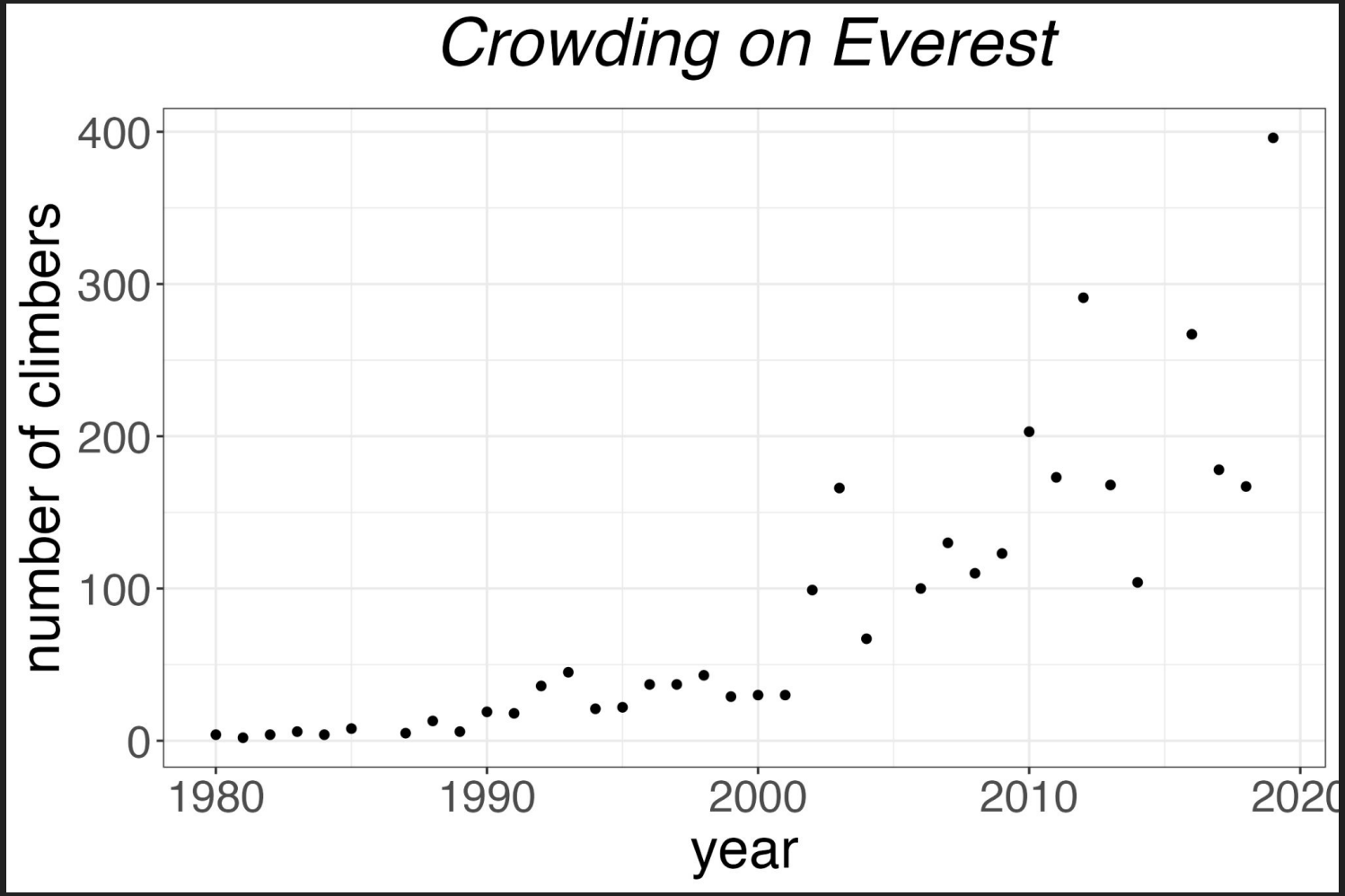
--Coding Demo--

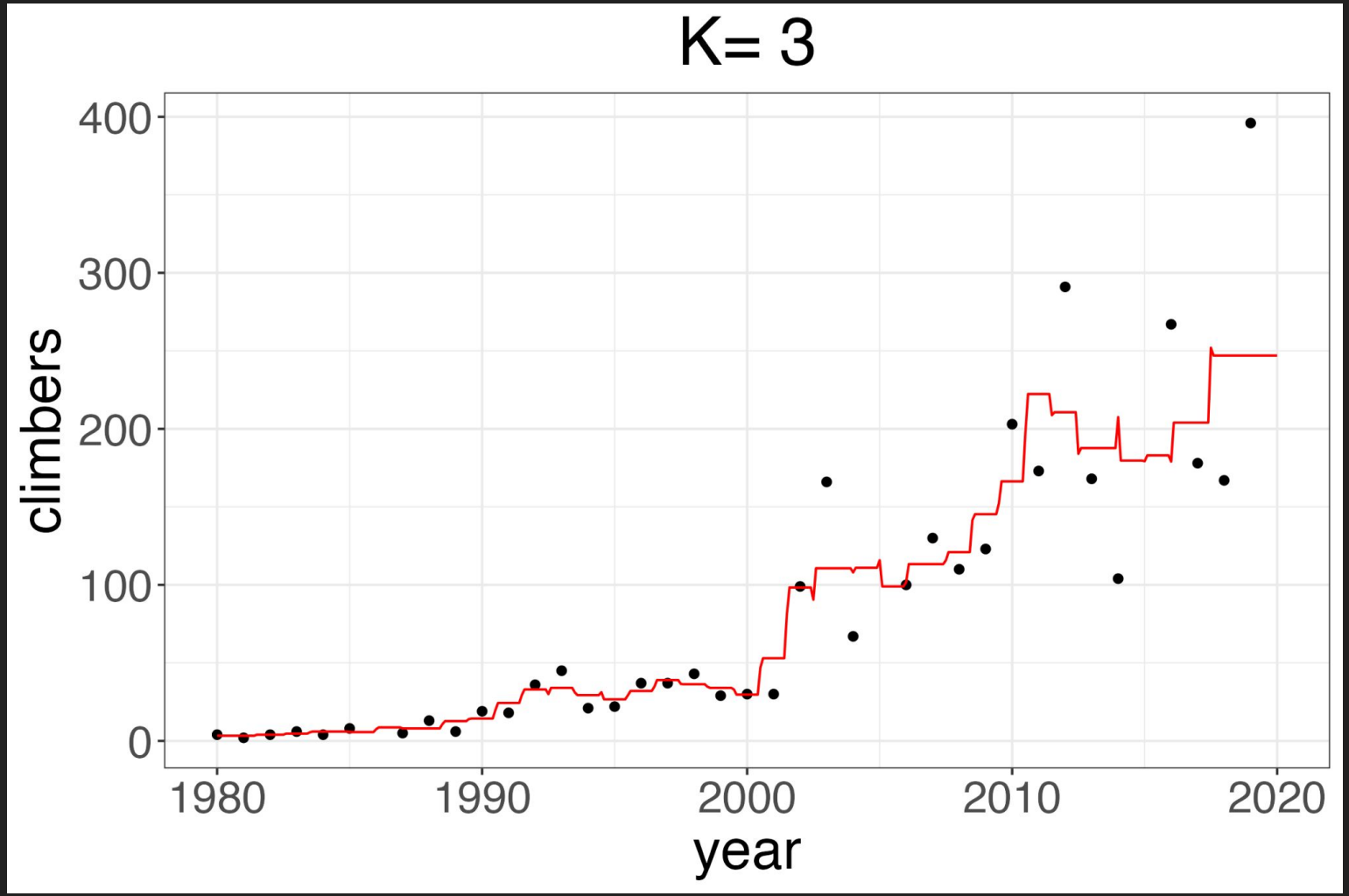
Results of K-fold CV

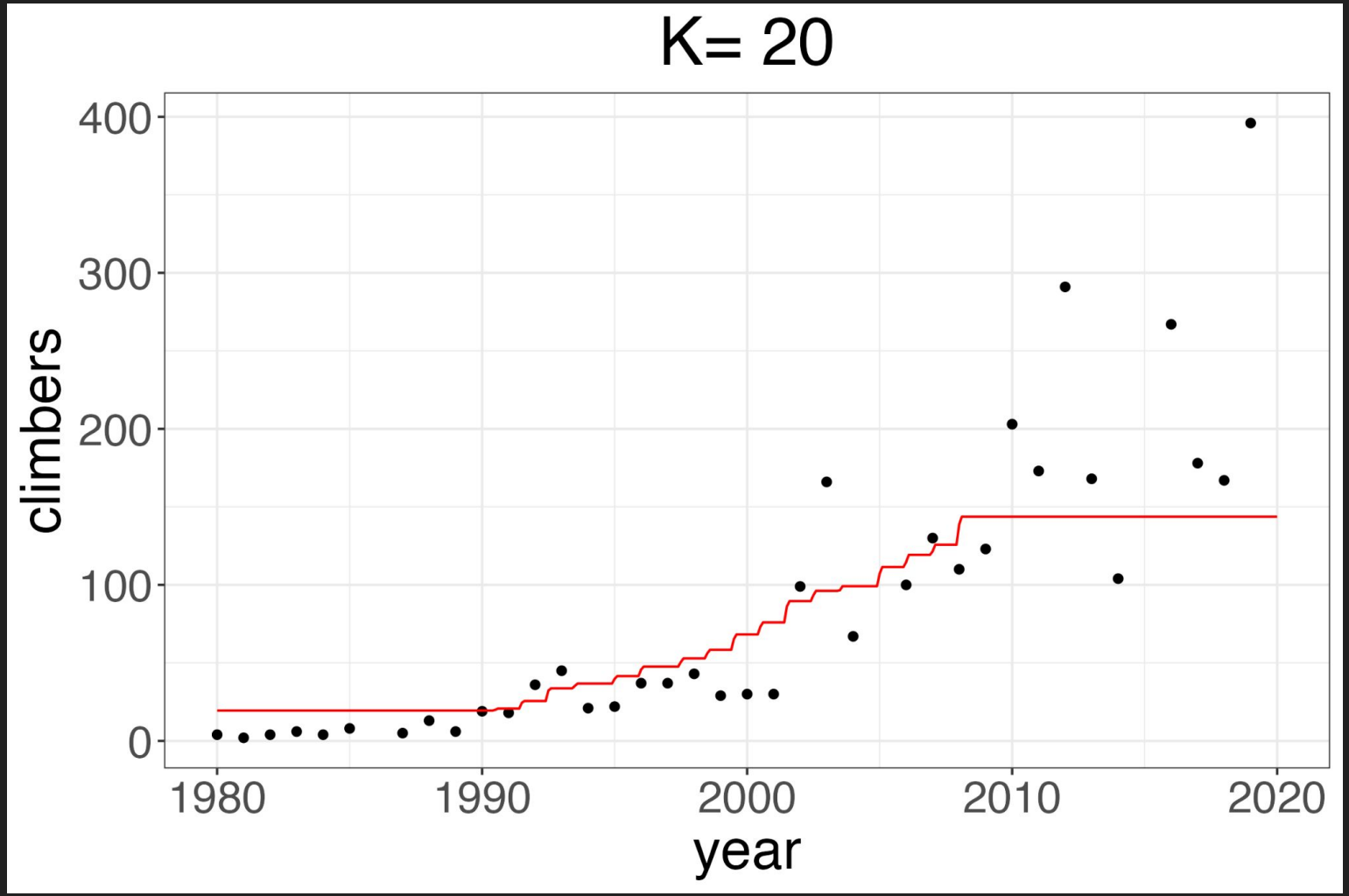

K= 3
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x

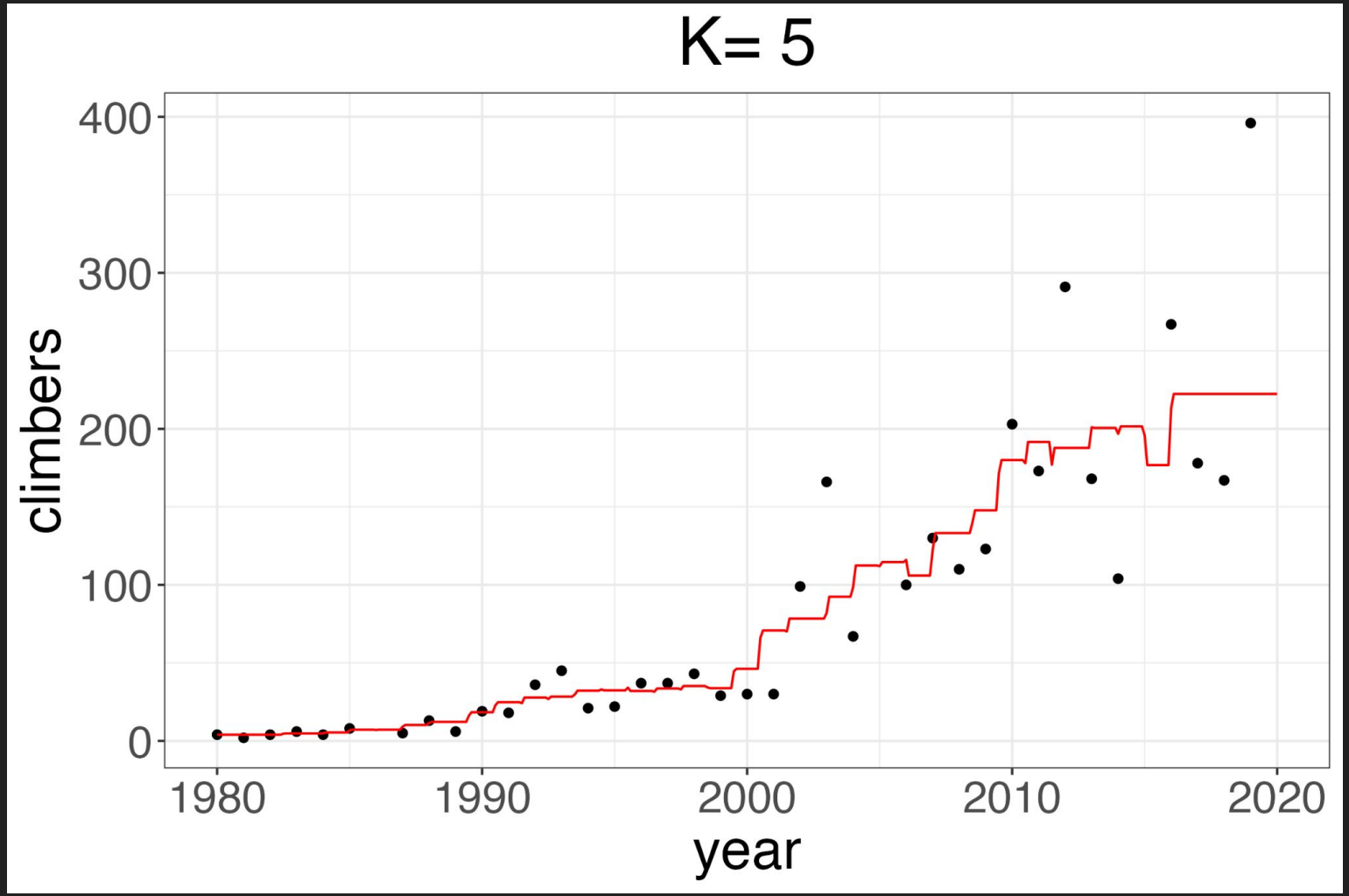

K= 10
14
12
10
8
g
0.0
2.5
5.0
7.5
10.0
x

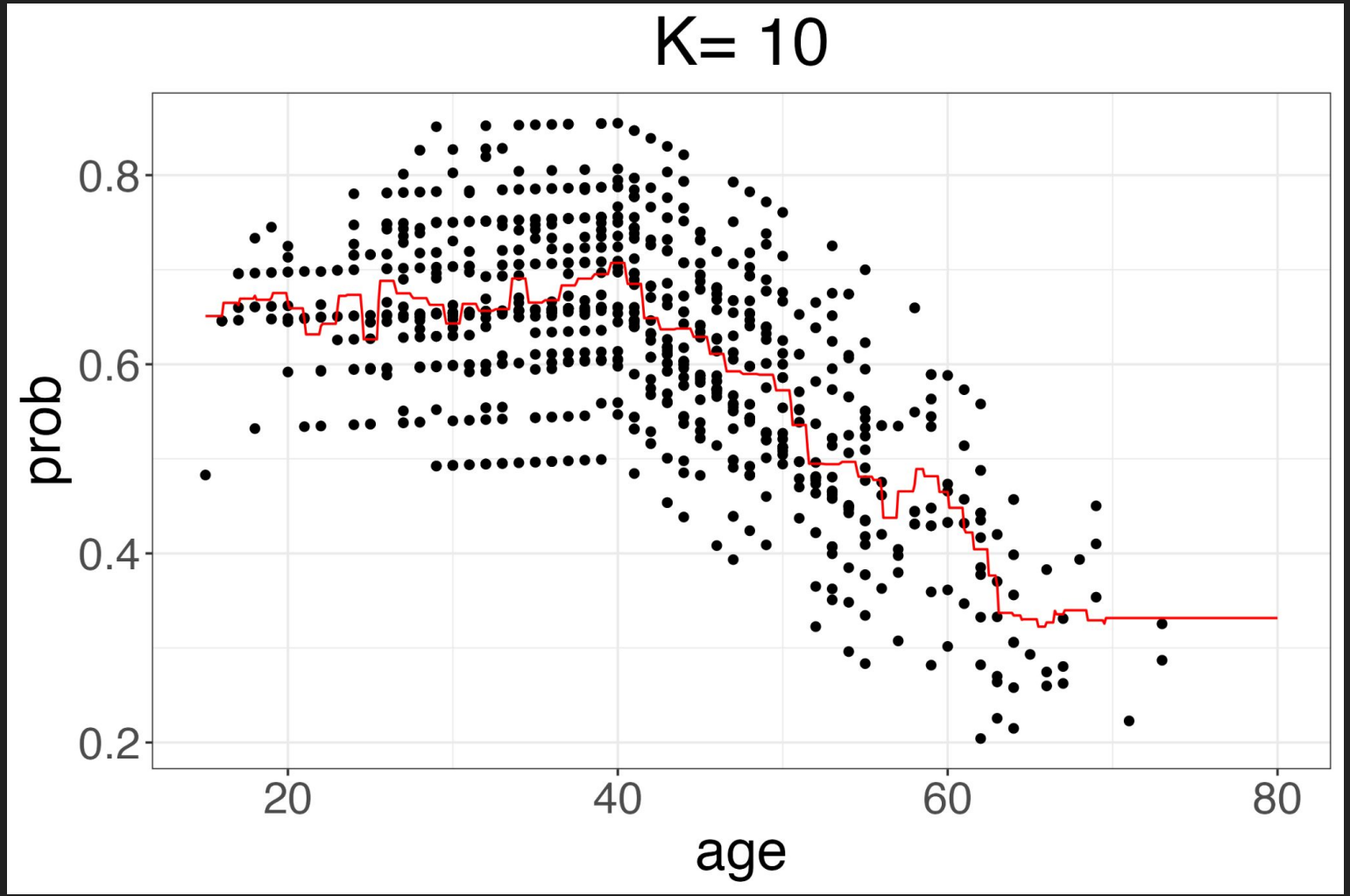
Back to Everest

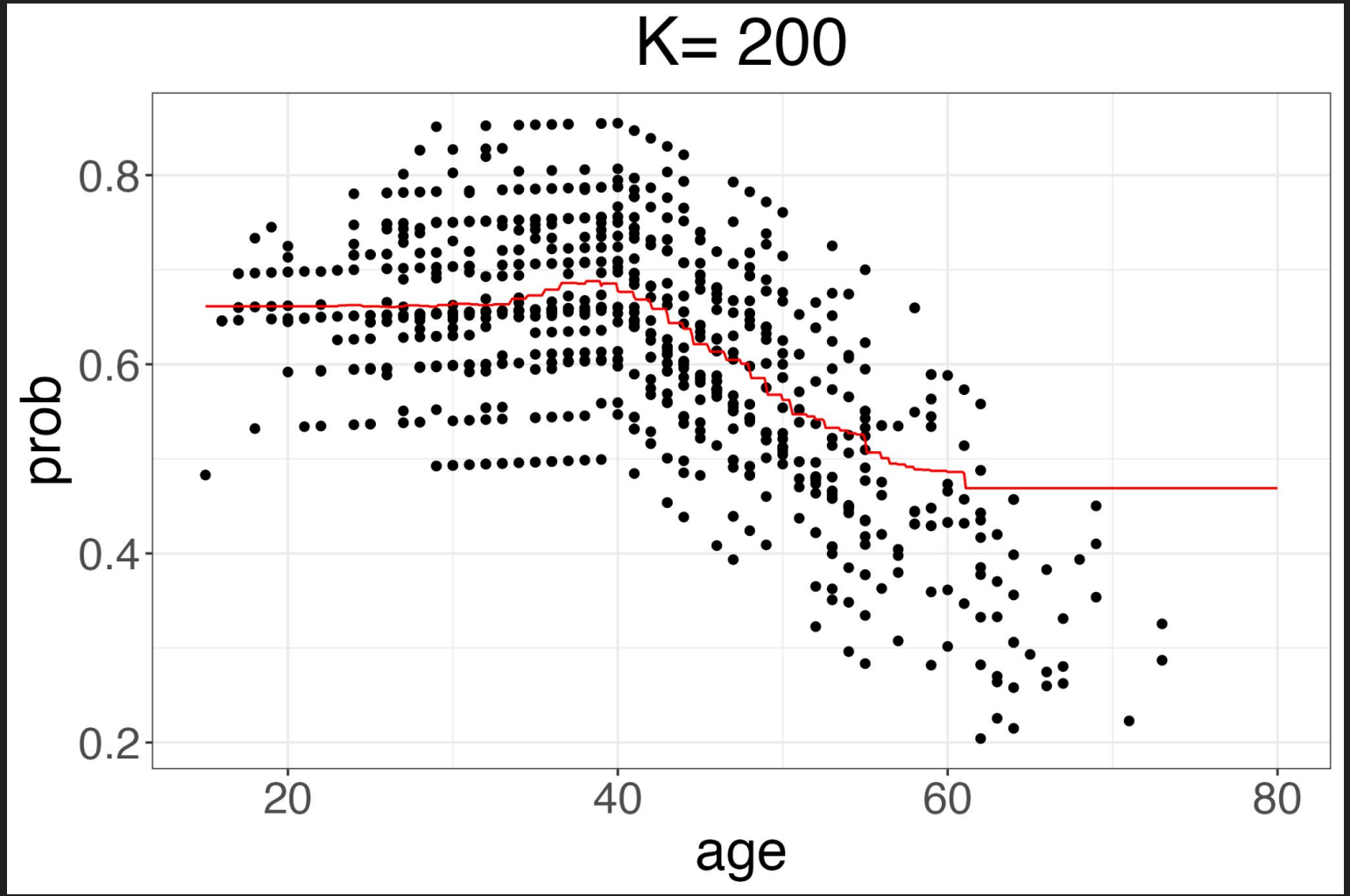

Crowding on Everest
number of climbers
0
100
200
300
400
1980
1990
2000
2010
2020
year

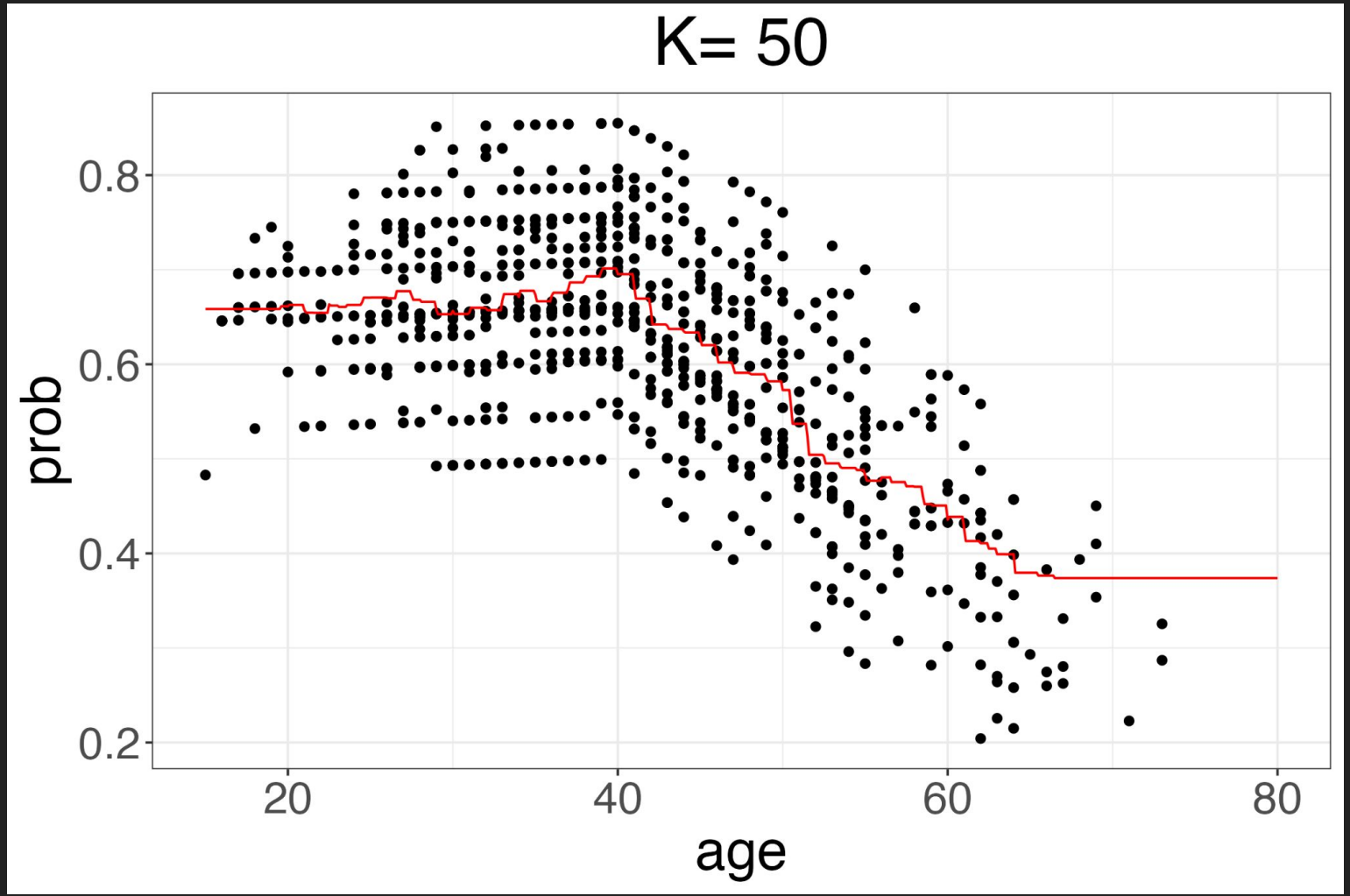

K= 3
400
300
200
100
0
climbers
1980
1990
2000
2010
2020
year


K= 20
400
300
200
100
0
climbers
1980
1990
2000
2010
2020
year


K= 5
400
300
200
100
0
climbers
1980
1990
2000
2010
2020
year


K= 10
prob
0.8
0.6
0.4
0.2
20
40
60
80
age


K= 200
prob
0.8
0.6
0.4
0.2
20
40
60
80
age


K= 50
prob
0.8
0.6
0.4
0.2
20
40
60
80
age

Some things to notice

- KNN regression is a weighted average

- Estimate minimizes a special loss function

Some things to notice

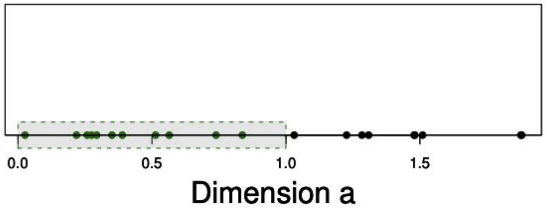
- Size of “neighborhood” changes with x
- Estimated curve is not continuous

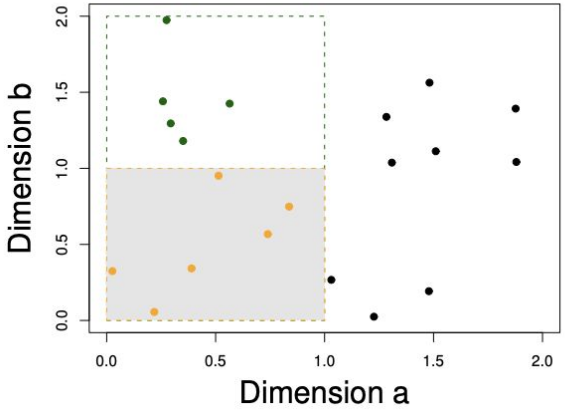
Strengths of KNN Reg

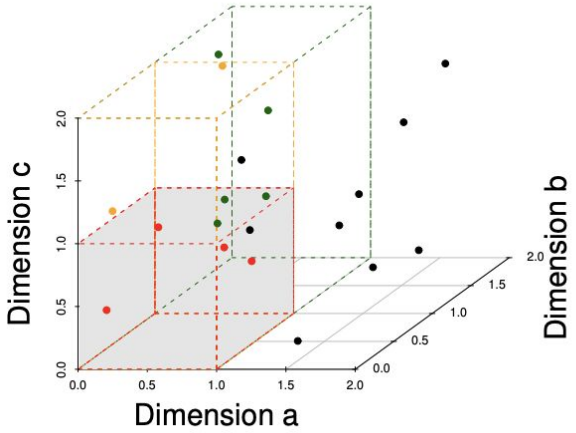
- Intuitive
- Local
- No parametric form assumed
- Easy to implement in software

Drawbacks of KNN Reg

- Sensitive to outliers
- Noticeable bias at boundaries (asymmetric neighbors)
- Less easy to interpret
- Many predictors → curse of dimensionality (dissimilar neighbors)







Source: Parsons et al (2004)

Other local nonparametric regression techniques

- Moving averages / kernel regression (e.g. Nadaraya and Watson 1964)

- Local linear / polynomial regression [e.g. LOESS] (Gram 1879)

Kahoot!