Introduction to Nonparamteric Regression: Kernel Estimation

Youmi Suk

School of Data Science, University of Virginia

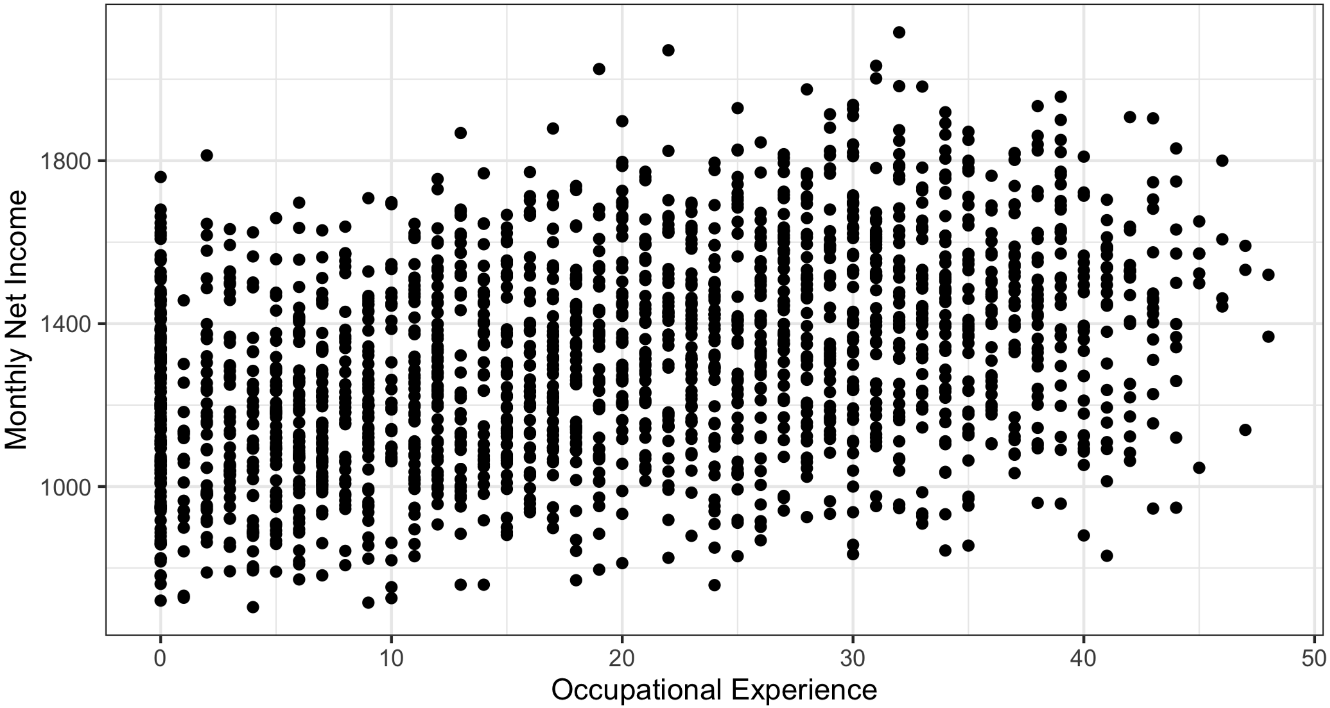
2. 2. 2022

Overview

1. Motivating Example
2. Conditional Distribution and Conditional Expectation
3. Binning
4. Local Averaging and Kernel Estimation

Motivating Example

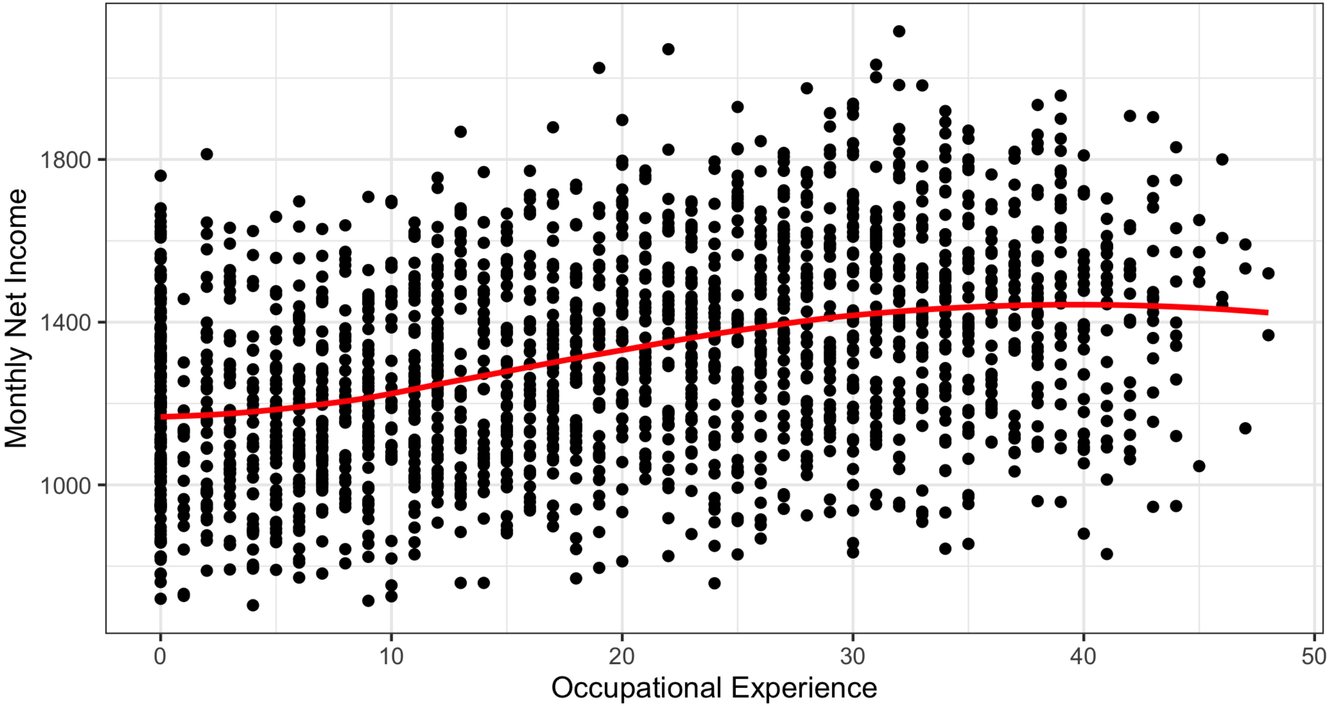
- Income and Occupational Experience:



- *Example*: simulated data of monthly net income (1922 observations)

Example: Income ~ Occup. Experience

- Scatterplot with a regression smoother:



What is Regression Analysis?

- Fox (2008):
    - Regression analysis examines the relationship between a quantitative response variable Y, and one or more quantitative explanatory variables $X_1, \ldots, X_k$.
    - Regression analysis traces the conditional distribution of Y—or some aspect of this distribution, such as its mean—as a function of explanatory variables $X_1, \ldots, X_k$, i.e., how does the distribution/mean of Y change with changes in the Xs?
- **Conditional distribution**

$$p(Y|x_1, x_2, \ldots, x_k) = g(x_1, x_2, \ldots, x_k)$$

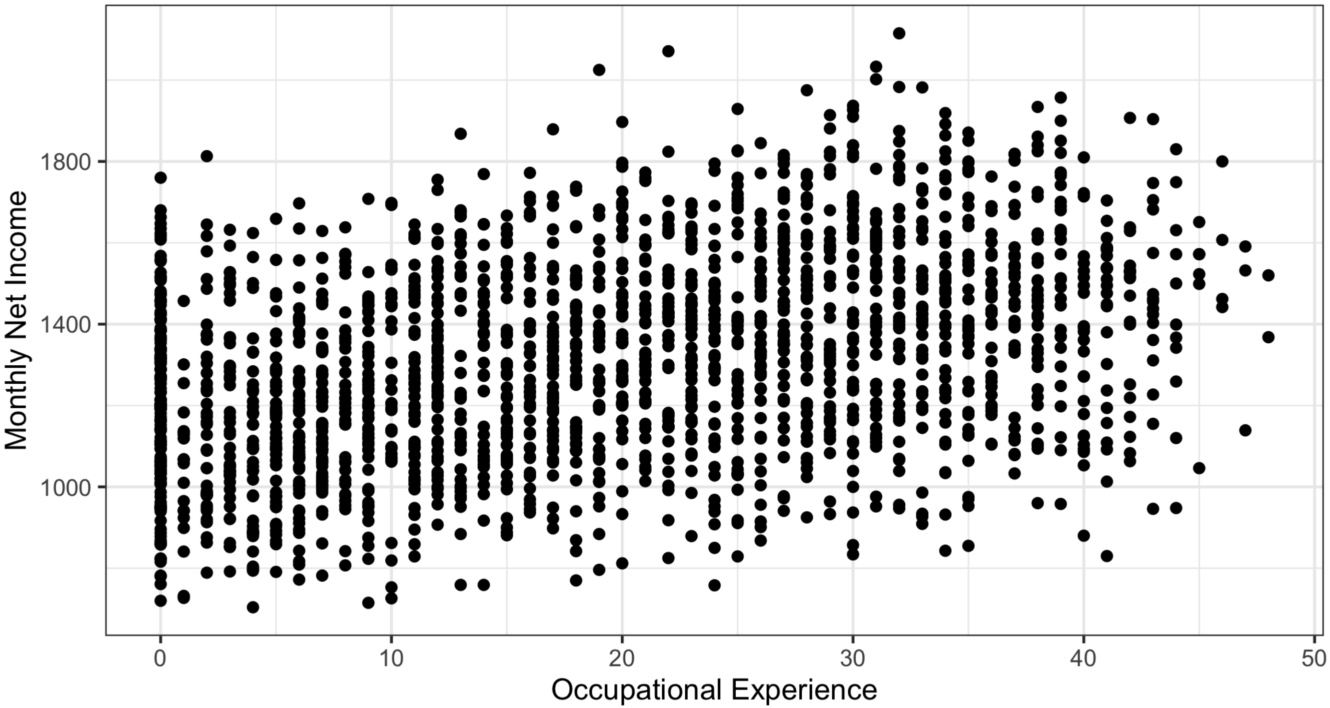
- **Conditional expectation**

$$E(Y|X_1, X_2, \ldots, X_k) = f(X_1, X_2, \ldots, X_k)$$

Regression Analysis: Example

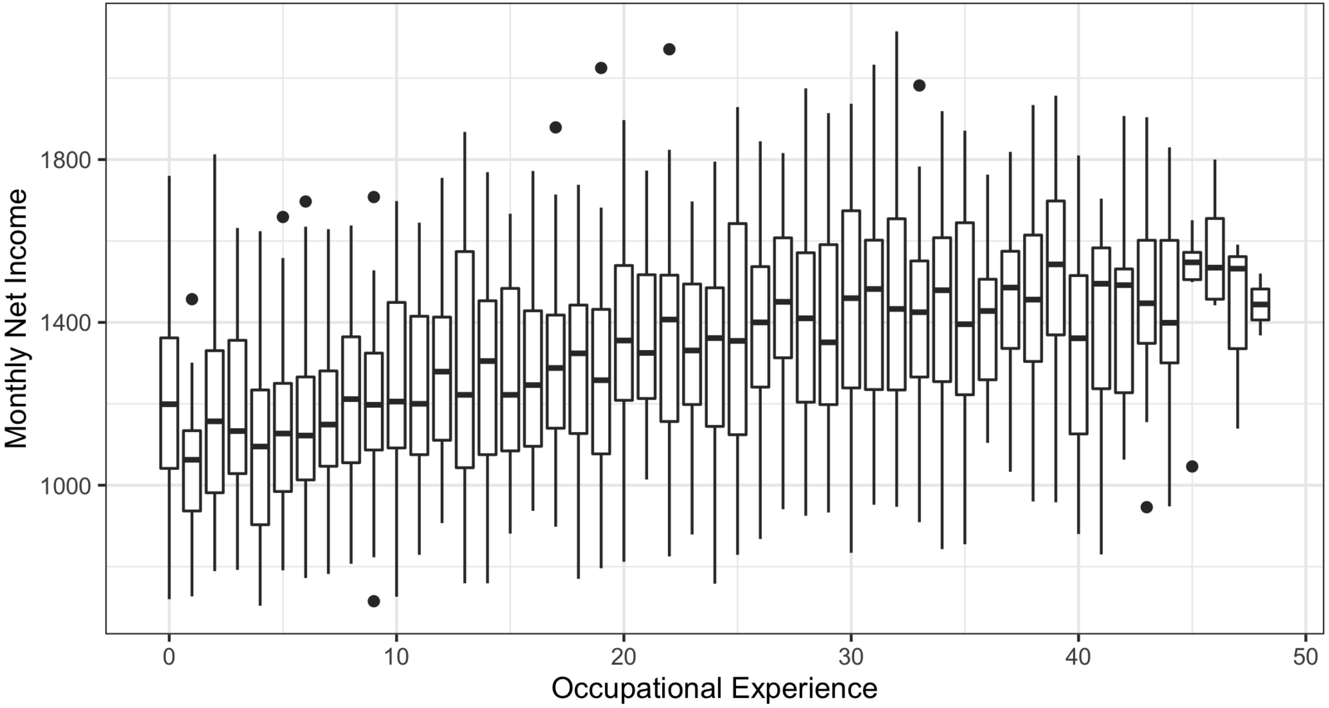
- Scatterplot: *Income* by *Occupational Experience*

```
(p <- ggplot(dat, aes(x=oexp, y=income)) + geom_point() + labs(x = '(
```



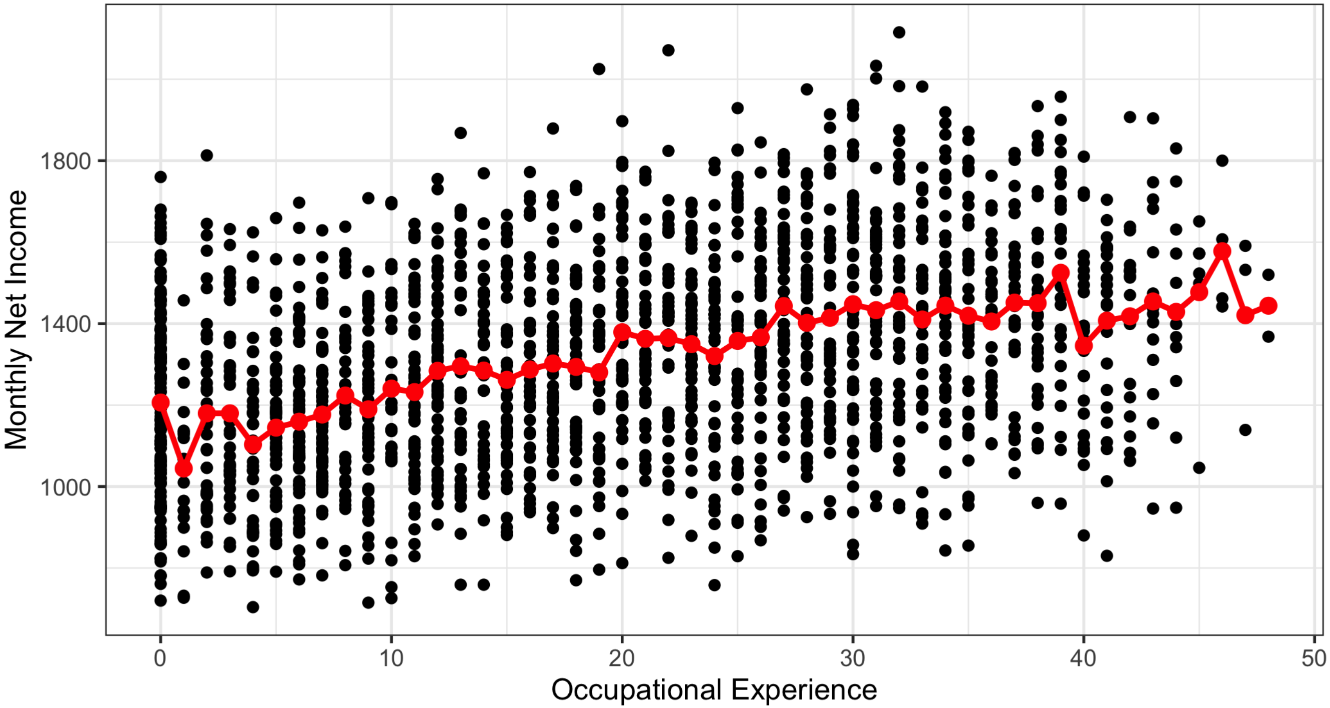
Conditional Distribution: Example

```
ggplot(dat, aes(x=oexp, y=income)) + geom_boxplot(aes(group=oexp)) +
```



Conditional Expectation: Example

```
p + stat_summary(fun = mean, geom = 'point', col = "red", size=2.5) +
    stat_summary(fun = mean, geom = 'line', col = "red", size=1)
```



Conditional Expectation Function (Path of Means)

The conditional expectation

$$E(Income|Occ.\,Experience = x)$$

tells us how the population average of *Income* changes as we move the conditioning variable (*Occ. Experience*) over all possible values. For every value of *Occupational Experience*, we might get a different average income. The collection of all such averages is called the conditional expectation function (CEF) and denoted as:

$$E(Income|Occ.\,Experience)$$

(without the specific value x for *Occ. Experience*)

More generally:

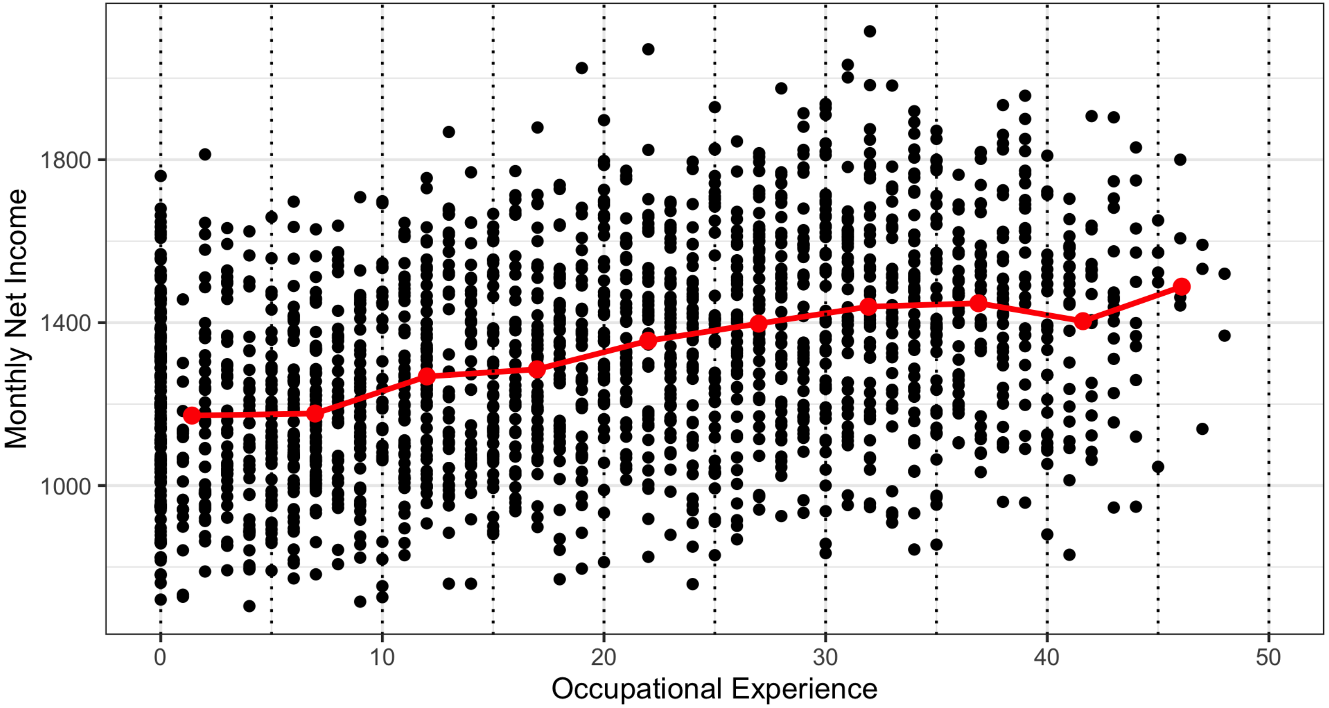
$$E(Y|X_1, X_2, \ldots, X_k)$$

Nonparametric Regression

- We investigate the conditional distribution/mean of a response variable without making any assumptions about the functional form or the shape of the distribution $\rightarrow$ nonparametric regression.
  - If an explanatory variable is discrete like age in years or educational level, we can estimate the conditional distribution/mean of Y directly.
  - If an explanatory variable is continuous (age measured in days, achievement scores) we don't have enough observations for each unique value; thus we may aggregate the explanatory variable into a large number of narrow bins and take the mean (or other statistic) value $\rightarrow$ binning

Binning

5-year bins: $C \in \{0-4, 5-9, \ldots, 45-49\}$ conditional mean given bins C



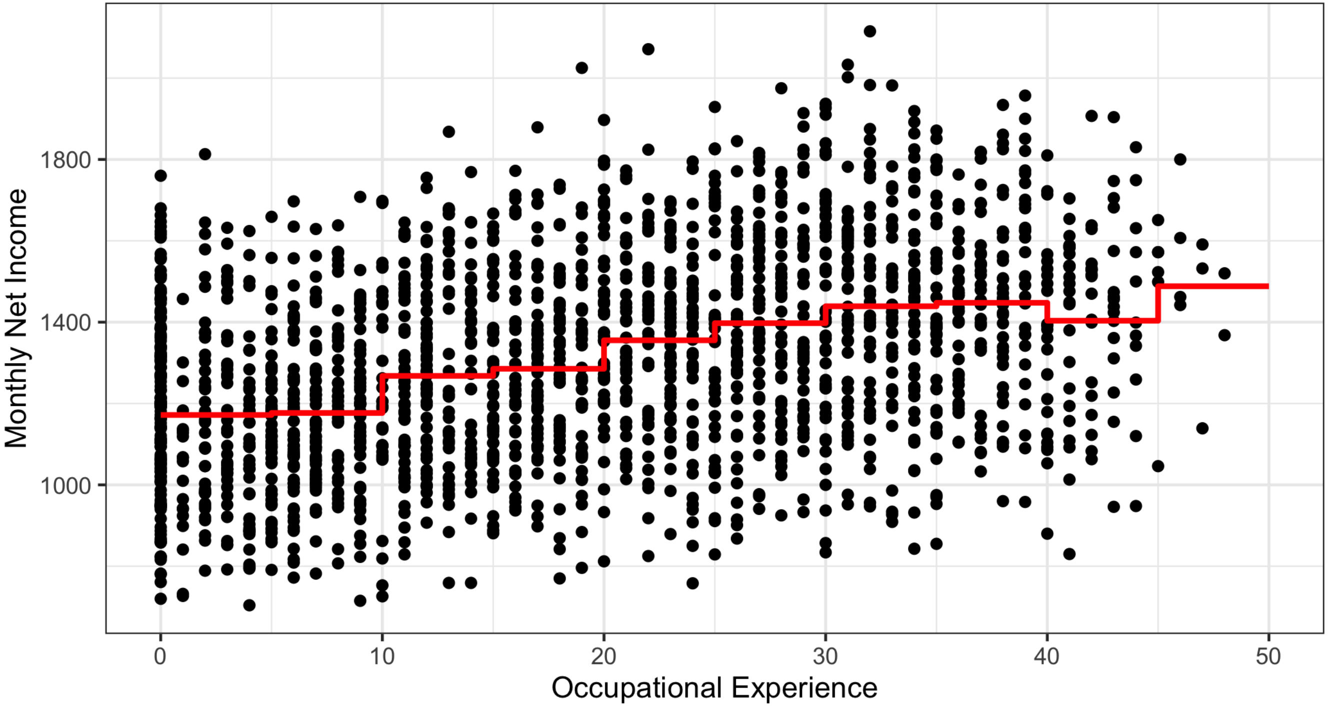
```r
# create 5-year age groups.
dat$oexpbins <- cut(dat$oexp, breaks = seq(0, 50, by = 5), right = F)

# compute the mean income and the mean oexp for each bin.
bin.stat <- dat %>% group_by(oexpbins) %>%
            summarize(m.income = mean(income), m.oexp = mean(oexp), 

# add points and lines for group means and add vertical lines at brea
p + geom_point(data=bin.stat, aes(x=m.oexp, y=m.income), col = "red",
    geom_line(data=bin.stat, aes(x=m.oexp, y=m.income), col = "red", 
    geom_vline(xintercept = seq(0, 50, by = 5), linetype="dotted", s
```

Binning

Conditional means as a step-function



Regression/Scatterplot Smoothers

- Binning is a very rough method and depends, of course, on the location and size of the bins.
- An alternative approach are so called regression or scatterplot smoothers which construct a separate bin (=window) for each unique x-value and then calculates the mean (or other statistic) from the data within the window. The window is typically defined by kernel weights.
  - Local averaging & kernel estimation
  - Local polynomial regression
  - Locally weighted regression (loess/lowess)

Local Averaging

- Local averaging calculates the mean (or other statistic) for a window centered at a specific value x_0.

- Procedure for local averaging:

1. Define value x_0 for which you want to calculate the mean

2. Construct a symmetric neighborhood around x_0 (=window); choose a window width using either

    - bandwidth (=half the window width in absolute measurement units of x) or

    - span (=a fraction of the data covered by the window, e.g., .3 = 30%).

3. For observations within the window, compute the mean of Y.

4. Do so for all other x-values

Local Averaging

- More formally, the weighted local average for a given value x_0 is given by

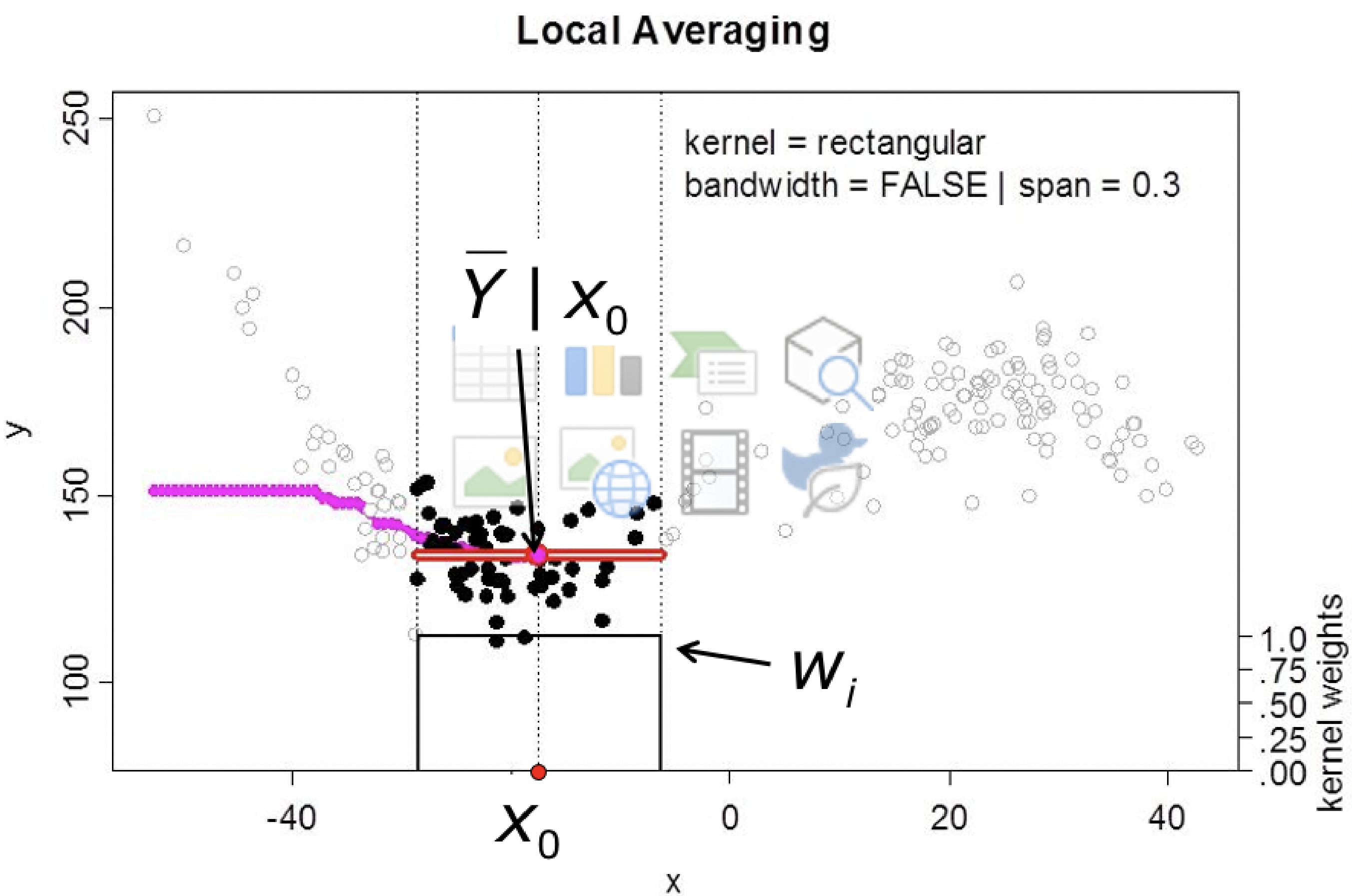
$$\overline{Y}|x_0 = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i}$$

where weight $w_i = 1$ if observation i lies within the window and $w_i = 0$ if it is outside the window:

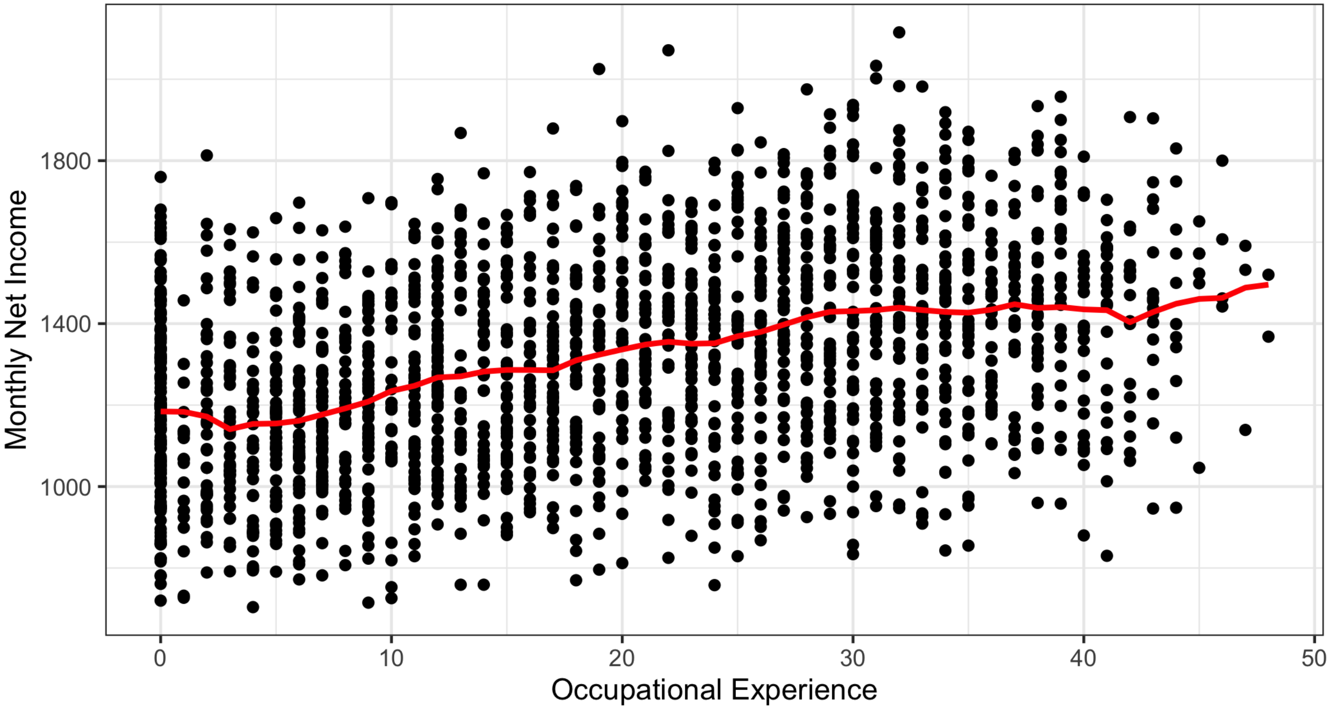
$$w_i = \begin{cases} 1 & \text{for } |(x_i - x_0)/h| < 1 \\ 0 & \text{for } |(x_i - x_0)/h| \geq 1 \end{cases}$$

h is the bandwidth of the window (half of the window width). Note that the weight function is identical to a rectangular kernel.

Local Averaging

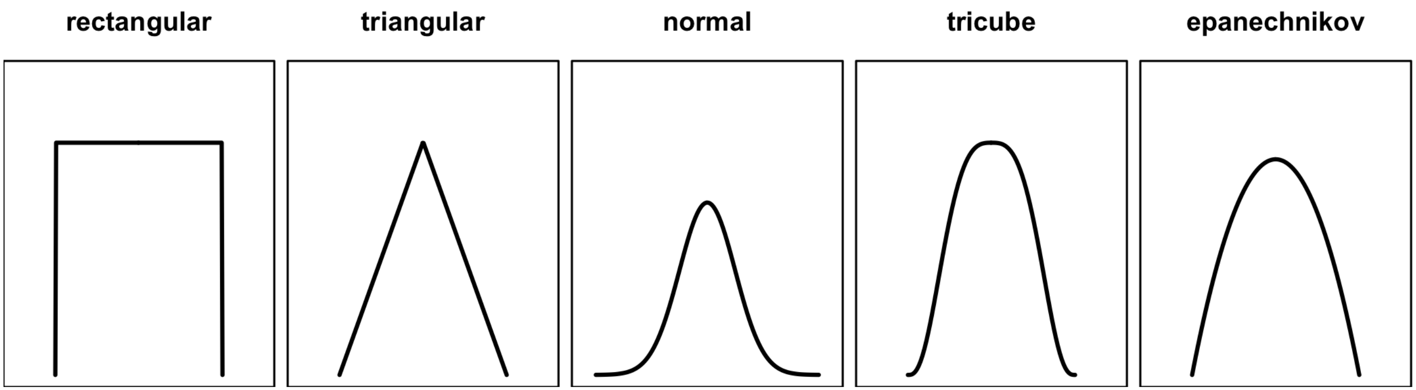


Local Averaging: Example

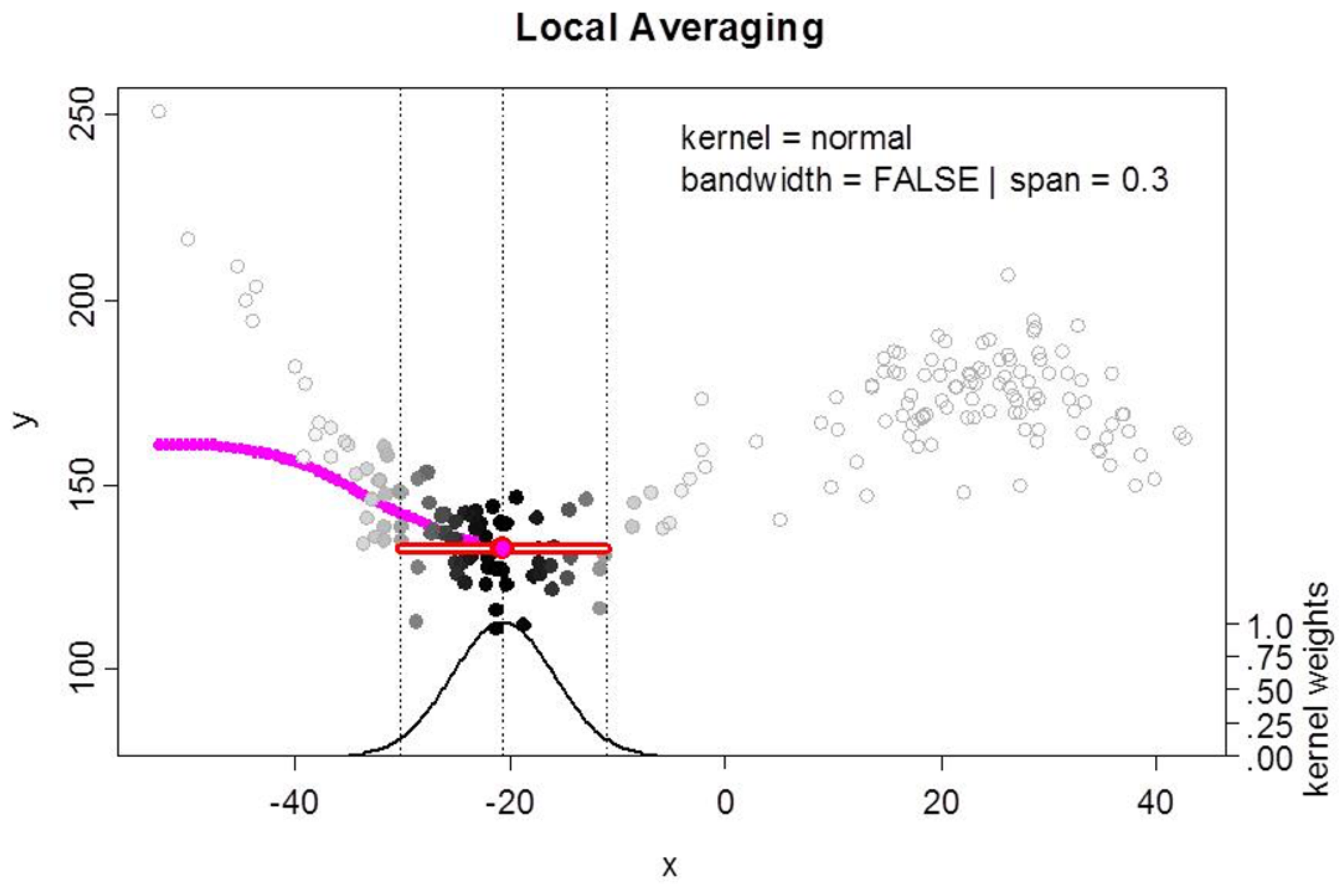


Kernel Estimation

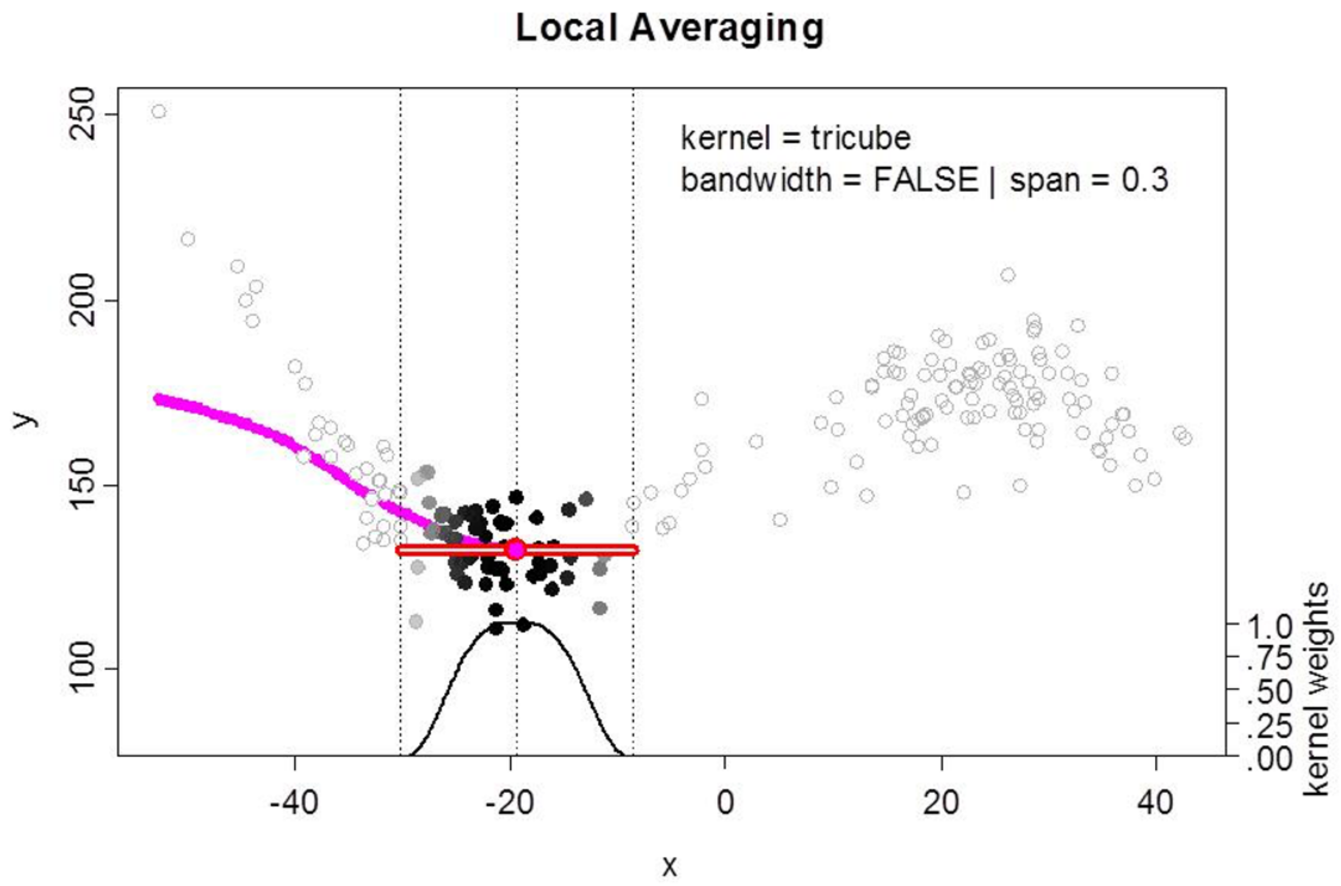
- We can do a local averaging by using different weighting functions, i.e., kernels
- Weights are defined by kernels (like in kernel density estimation), e.g.
  - Rectangular (uniform) kernel: same weight for all observations within the window = local averaging
  - Triangular, normal, tricube, Epanechnikov kernel: observations closer to the x_0 under consideration get more weight than observations further away



Kernel Estimation: Normal Kernel

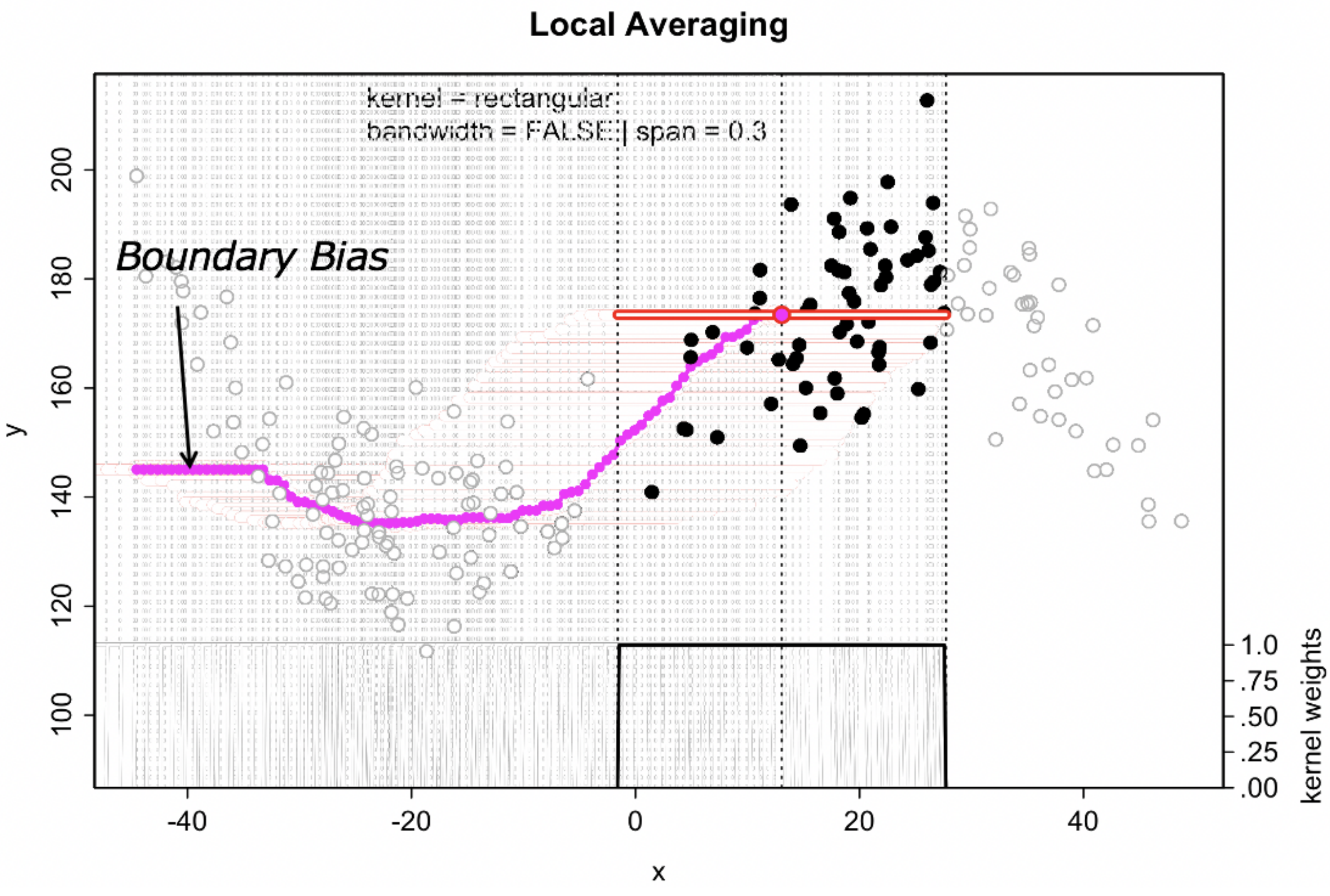


Kernel Estimation: Tricube Kernel



R Demonstration

Boundary Bias



Regression/Scatterplot Smoothers

- Binning is a very rough method and depends, of course, on the location and size of the bins.
- An alternative approach are so called regression or scatterplot smoothers which construct a separate bin (=window) for each unique x-value and then calculates the mean (or other statistic) from the data within the window. The window is typically defined by kernel weights.
  - Local averaging & kernel estimation
  - Local polynomial regression
  - Locally weighted regression (loess/lowess)