



DATA VISUALIZATION IN R WITH LATTICE

Welcome to the course

Deepayan Sarkar

Associate Professor, Indian Statistical Institute

Data visualization

- Exploratory data analysis
- Presenting or reporting results

Data visualization frameworks in R

- Base R graphics
- lattice based on "Trellis graphics" (Cleveland)
- ggplot2 based on "Grammar of Graphics" (Wilkinson)

Number of packages depending on these (March 2017):

	graphics	lattice	ggplot2
CRAN	5612	3654	1566
CRAN+BioC	7889	4858	2038

The USCancerRates dataset

- Age-adjusted death rates due to cancer (per 100,000)
- Separately for males and females
- County-level data for 1999-2003
- Available in the latticeExtra package

The USCancerRates dataset

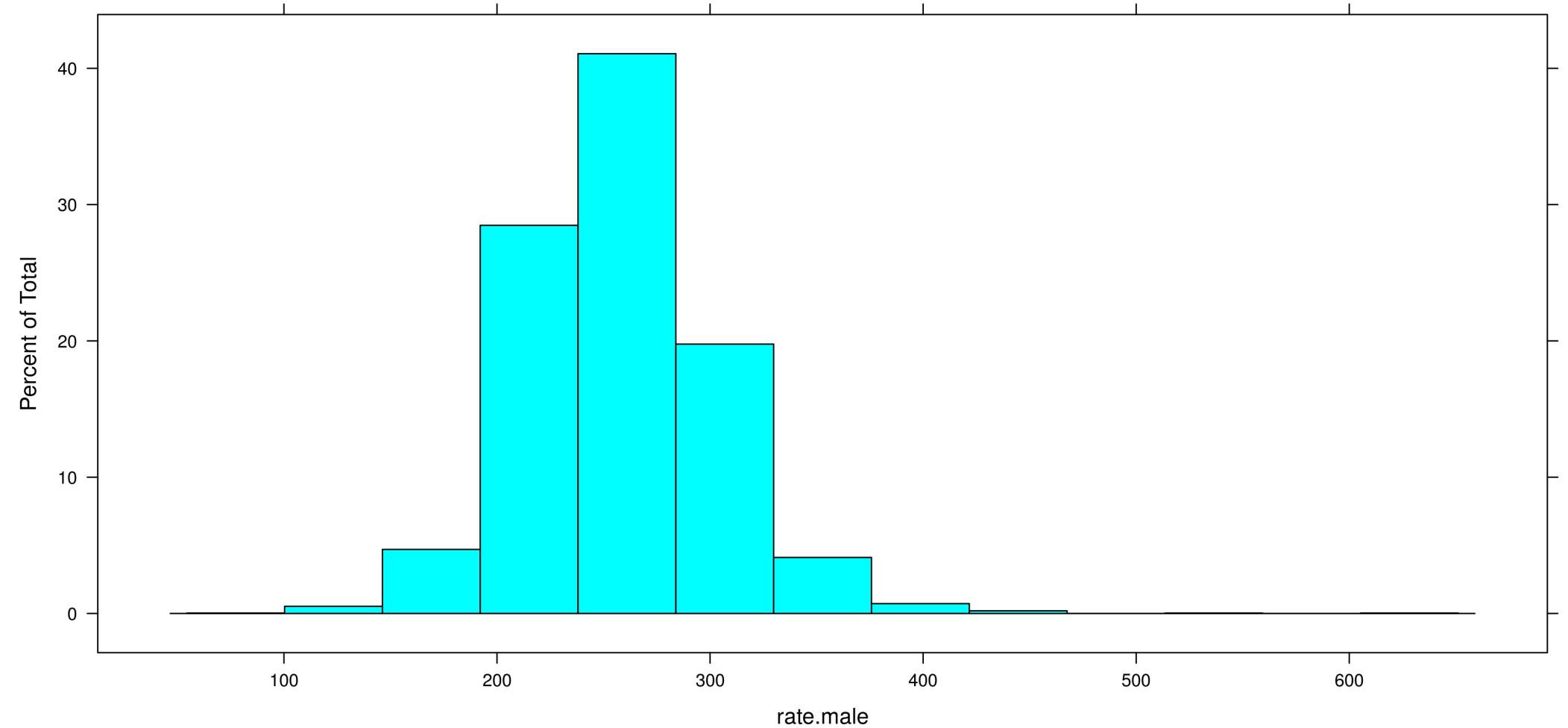
```
> library(lattice)
> data(USCancerRates, package = "latticeExtra")
> str(USCancerRates)
```

'data.frame': 3041 obs. of 8 variables:

\$ rate.male	: num 364 346 341 336 330 ...
\$ LCL95.male	: num 311 274 304 289 293 ...
\$ UCL95.male	: num 423 431 381 389 371 ...
\$ rate.female	: num 151 140 182 185 172 ...
\$ LCL95.female	: num 124 103 161 157 151 ...
\$ UCL95.female	: num 184 190 206 218 195 ...
\$ state	: Factor w/ 49 levels "Alabama", "Alaska", ... : 1 1 1 1 1 ...
\$ county	: chr [1:3041] "Pickens County" "Bullock County" ...

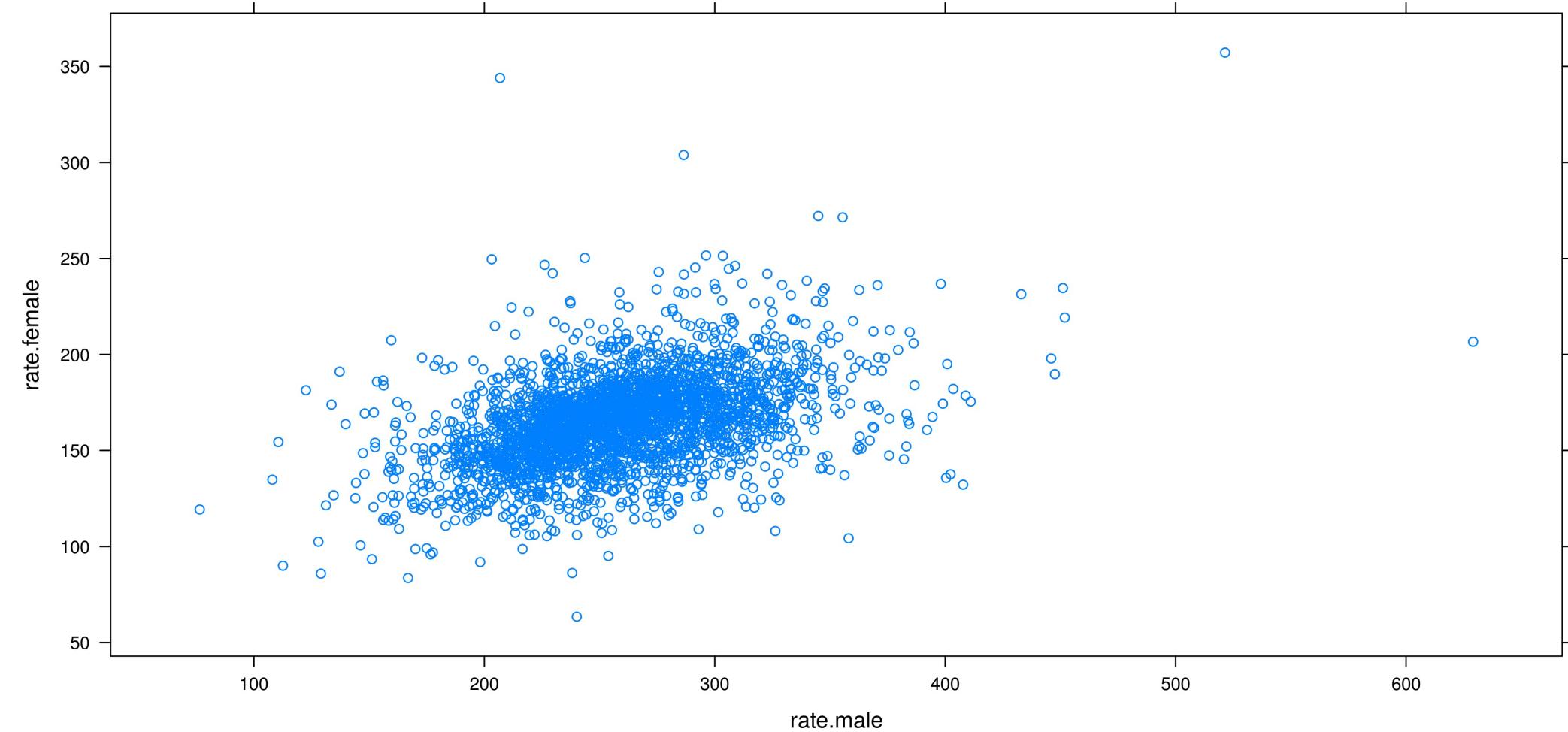
A histogram

```
> histogram(~ rate.male, data = USCancerRates)
```



A scatter plot

```
> xyplot(rate.female ~ rate.male, data = USCancerRates)
```



The formula

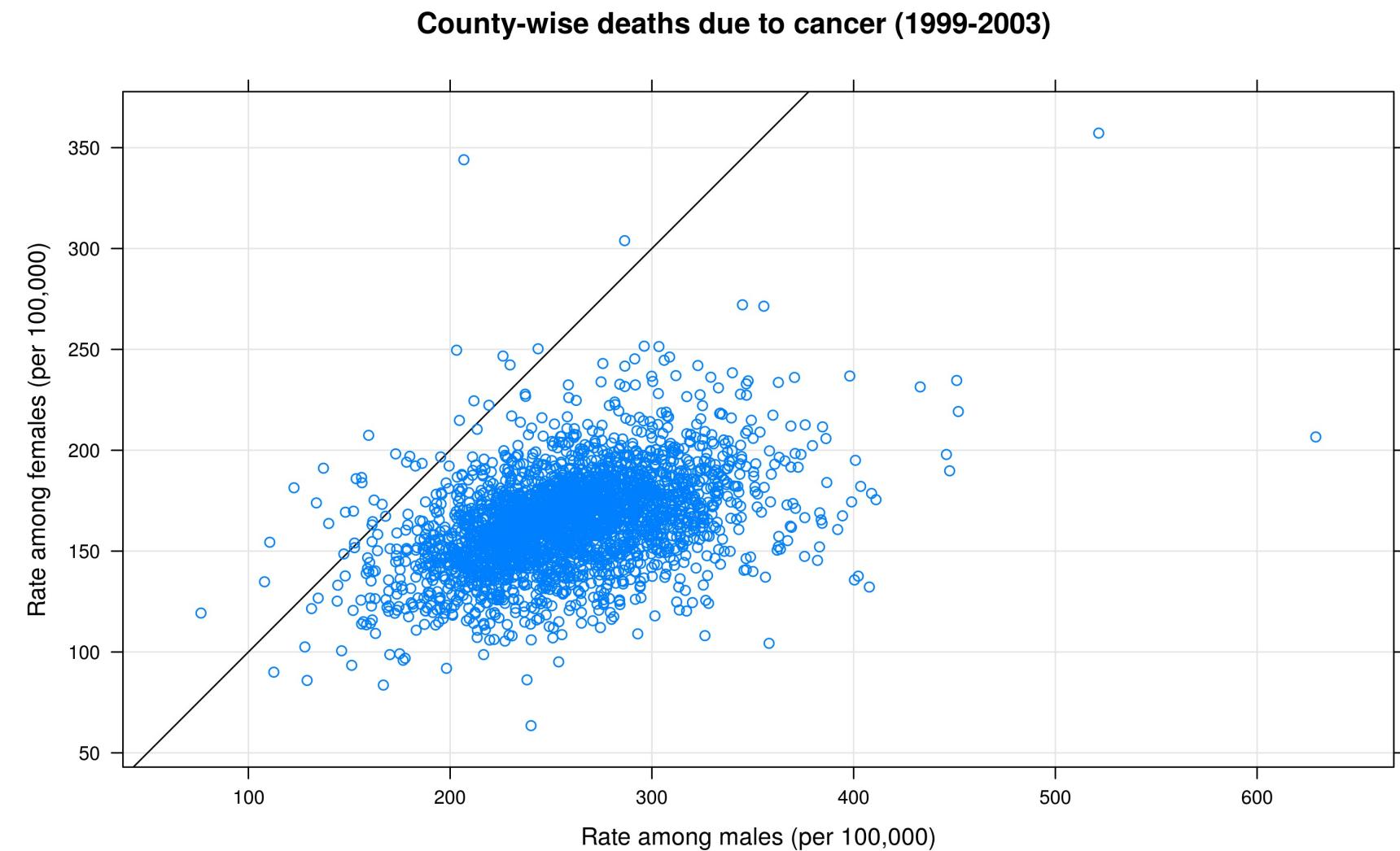
```
histogram(~ rate.male, data = USCancerRates)
xyplot(rate.female ~ rate.male, data = USCancerRates)
```

- $\sim x$ in `histogram()`: x plotted on x -axis
- $y \sim x$ in `xyplot()`:
 - x plotted on x -axis
 - y plotted on y -axis

Similar to modeling calls

```
lm(rate.female ~ rate.male, data = USCancerRates)
```

A version for presentation





DATA VISUALIZATION IN R WITH LATTICE

Let's practice!



DATA VISUALIZATION IN R WITH LATTICE

Optional arguments

Deepayan Sarkar

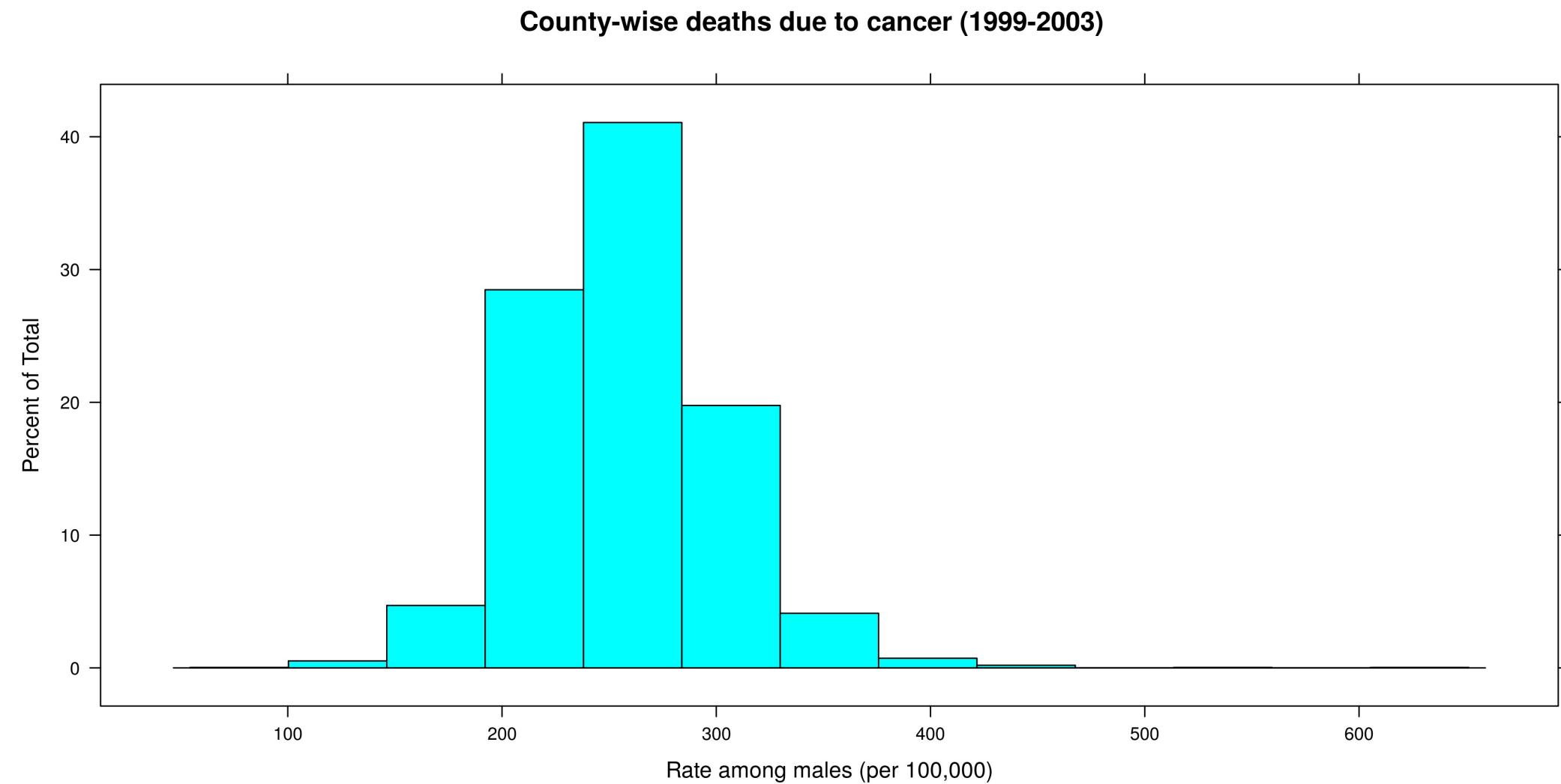
Associate Professor, Indian Statistical Institute

Arguments in lattice functions

- Mandatory:
 - x: formula (first argument, usually not named)
 - data: dataset containing variables
- Optional:
 - Some apply to all functions
 - Some are specific to particular functions

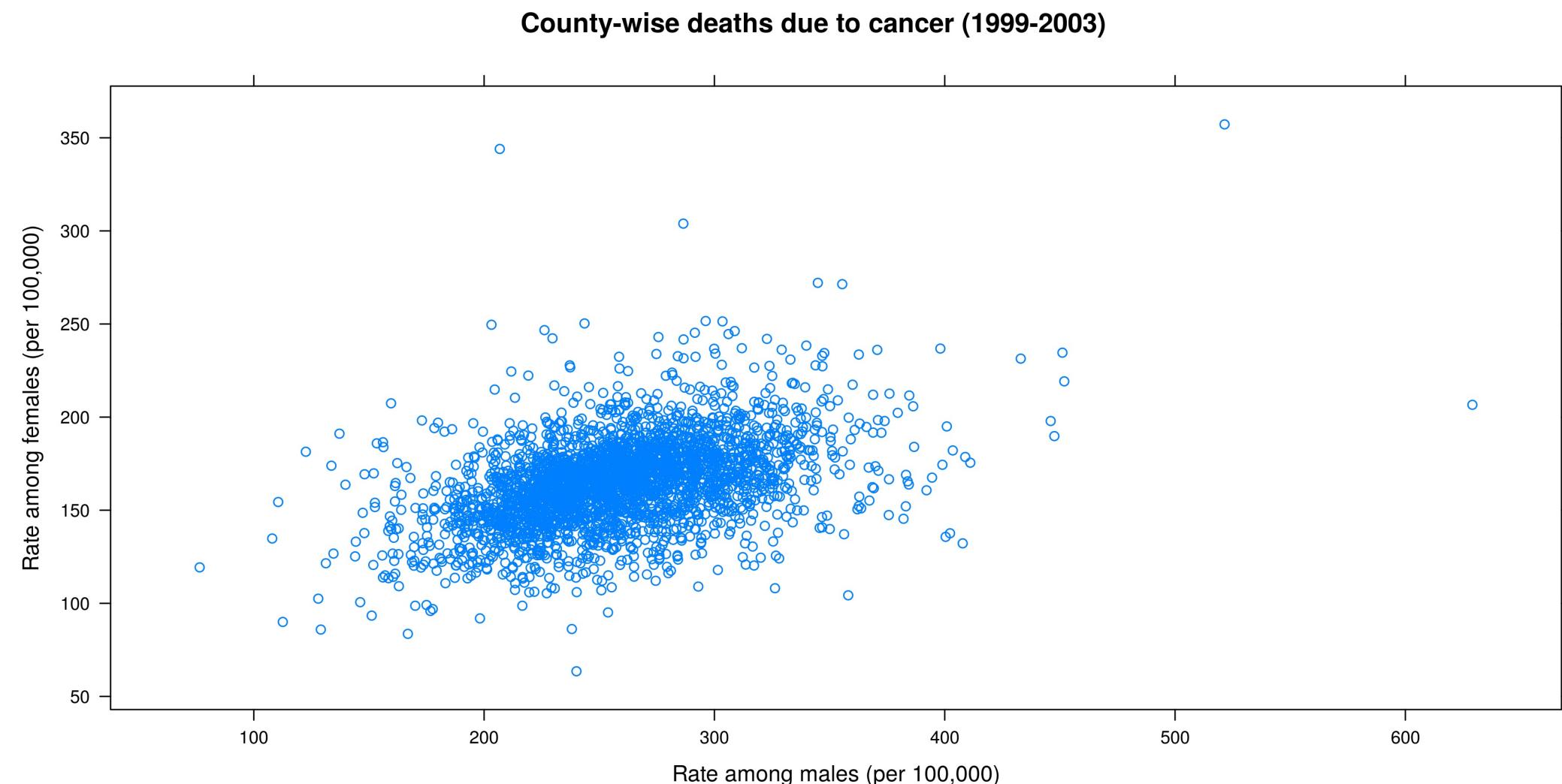
Common arguments: main, xlab, ylab

```
> histogram(~ rate.male, data = USCancerRates,  
+           main = "County-wise deaths due to cancer (1999-2003)",  
+           xlab = "Rate among males (per 100,000)")
```



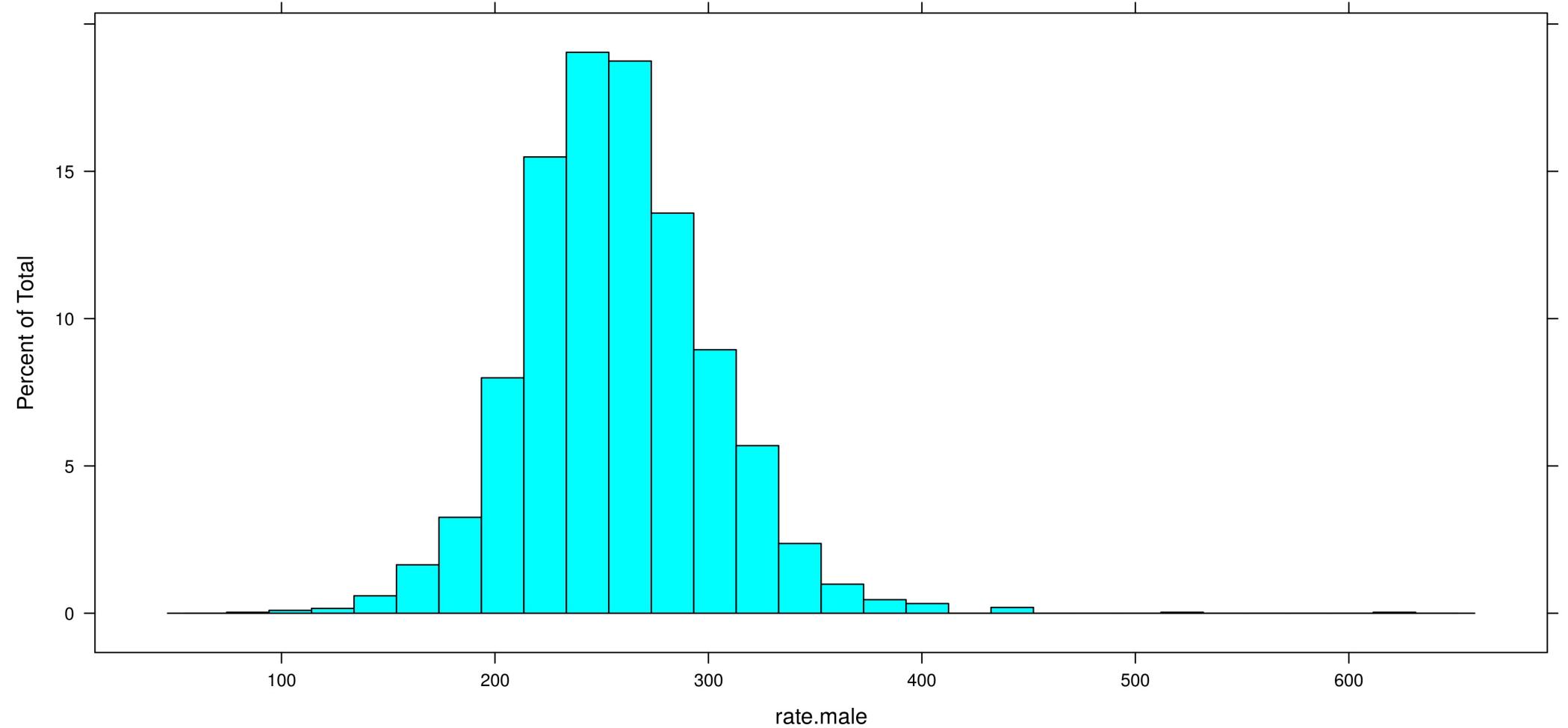
Common arguments: main, xlab, ylab

```
> xyplot(rate.female ~ rate.male, data = USCancerRates,  
+         main = "County-wise deaths due to cancer (1999-2003)",  
+         xlab = "Rate among males (per 100,000)",  
+         ylab = "Rate among females (per 100,000)")
```



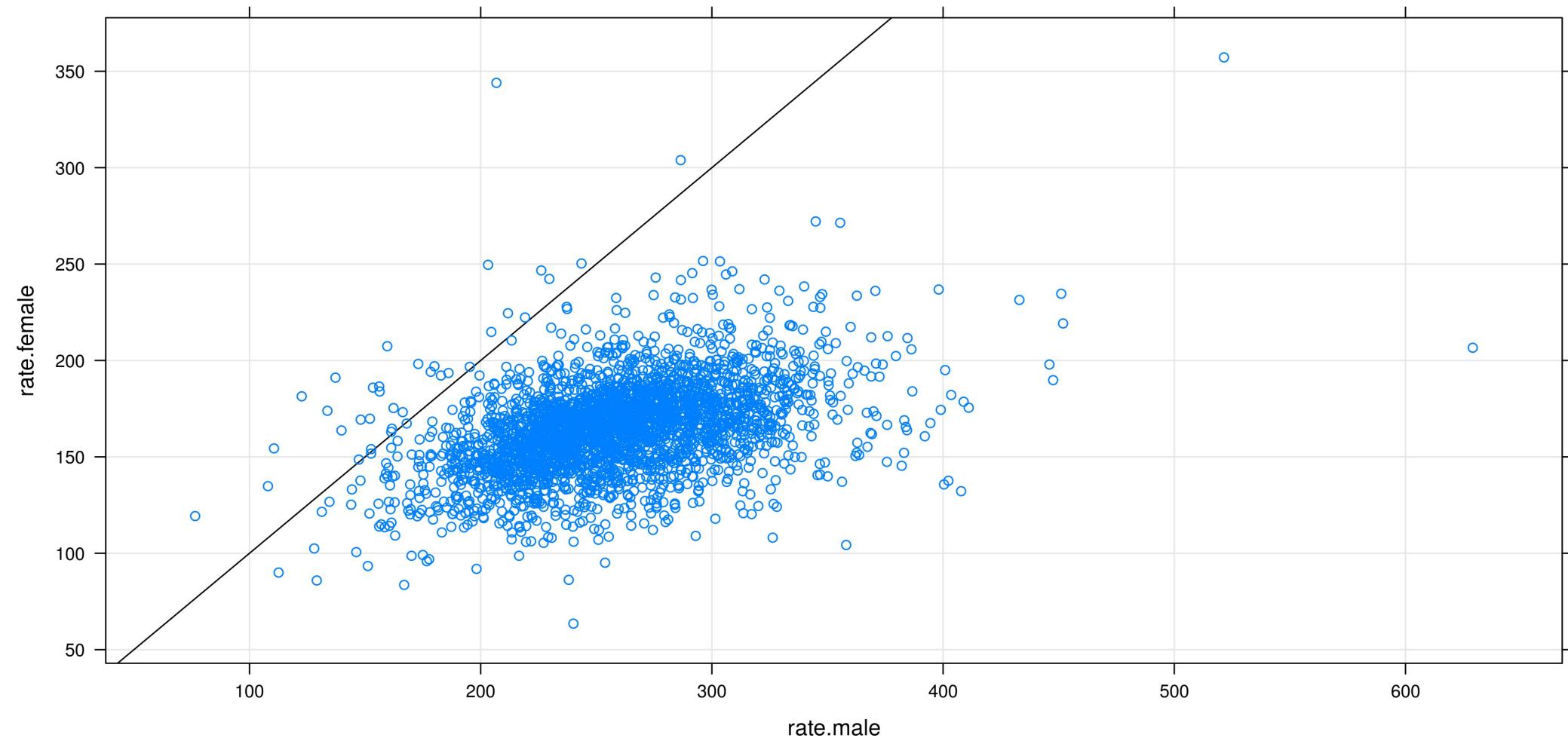
Arguments of histogram()

```
> histogram(~ rate.male, data = USCancerRates, nint = 30)
```



Arguments of xyplot()

```
> xyplot(rate.female ~ rate.male, data = USCancerRates,  
+         grid = TRUE, abline = c(0, 1))
```





DATA VISUALIZATION IN R WITH LATTICE

Let's practice!



DATA VISUALIZATION IN R WITH LATTICE

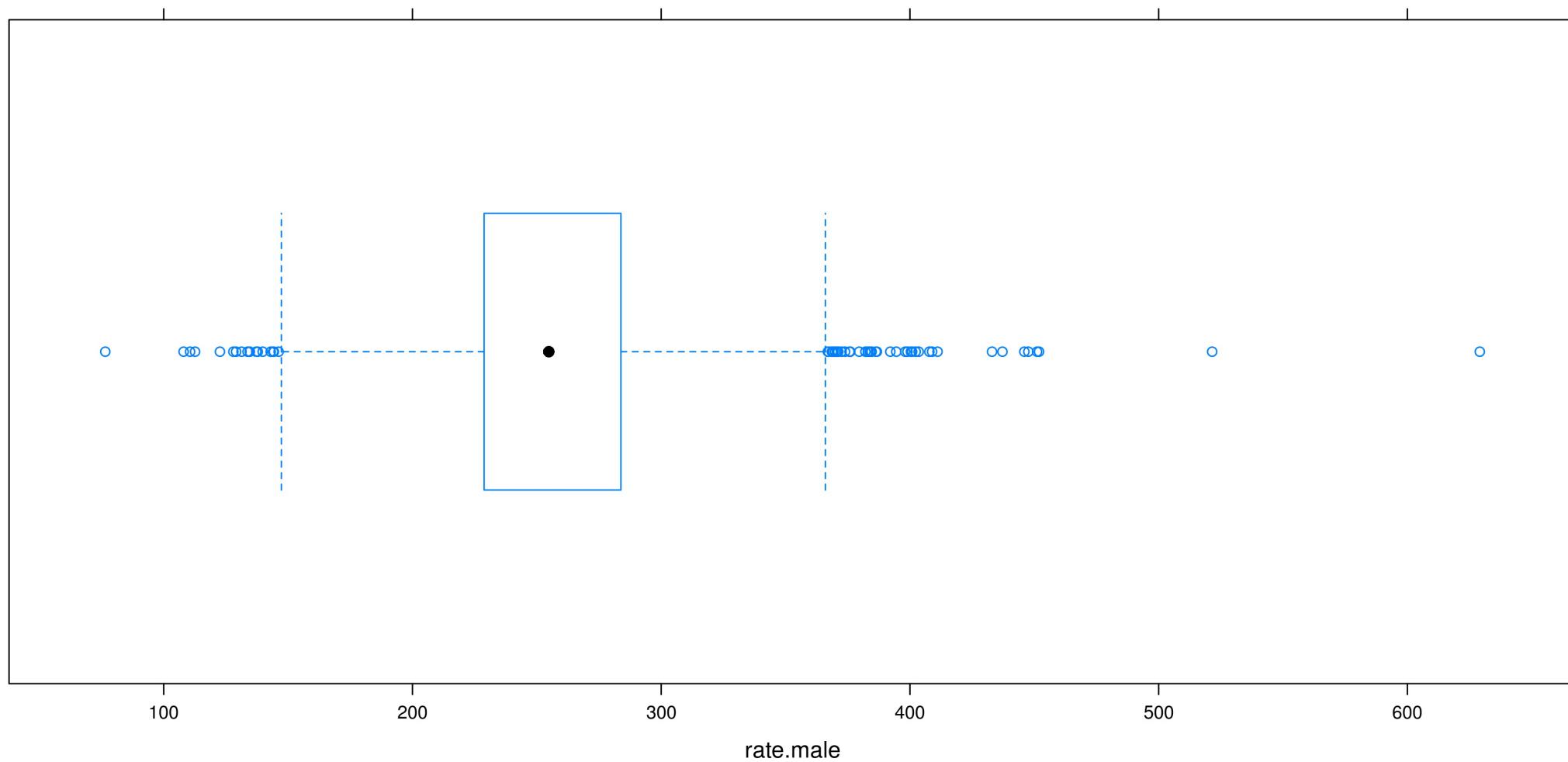
Box and whisker plots and reordering levels

Deepayan Sarkar

Associate Professor, Indian Statistical Institute

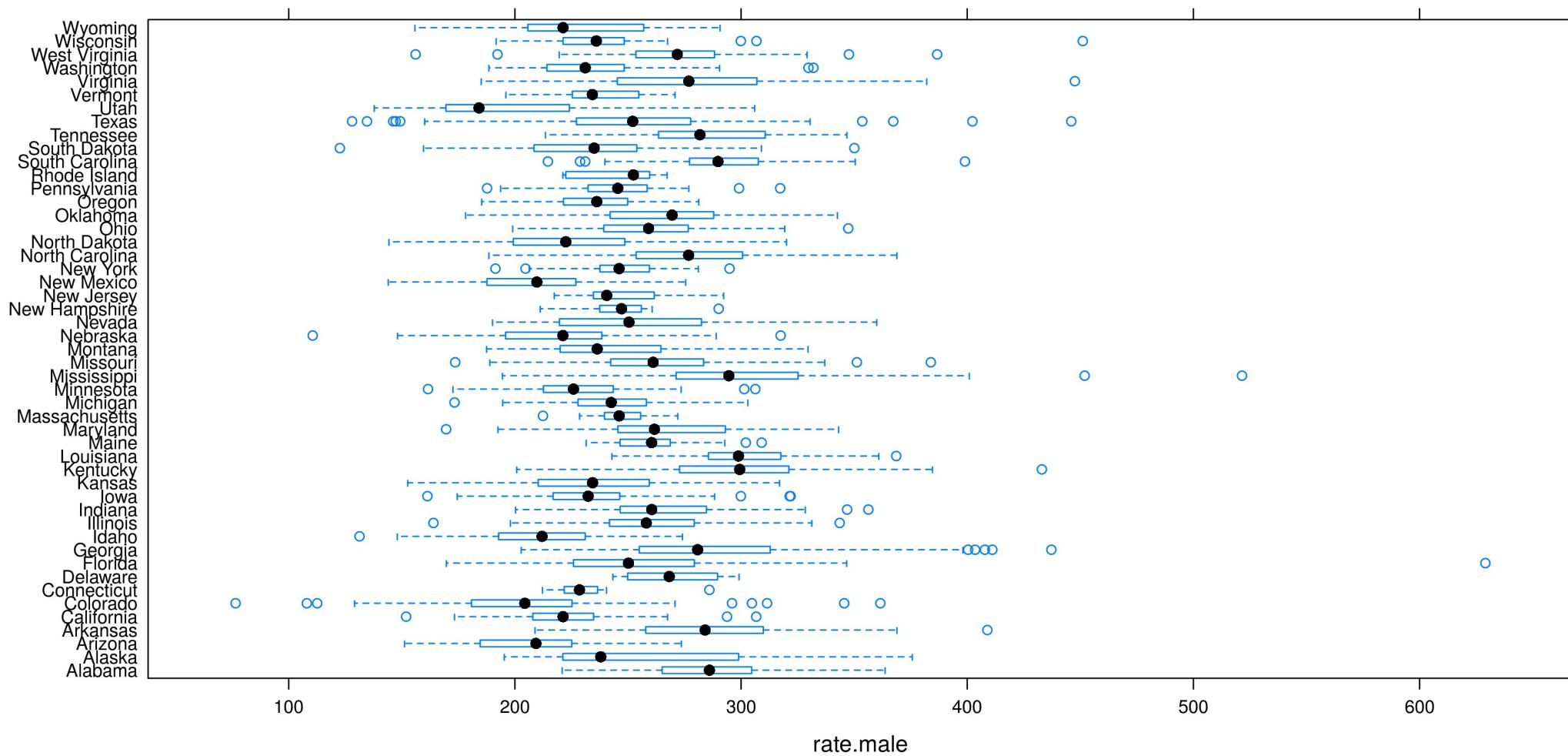
Box and whisker plots: bwplot()

```
> bwplot(~ rate.male, data = USCancerRates)
```



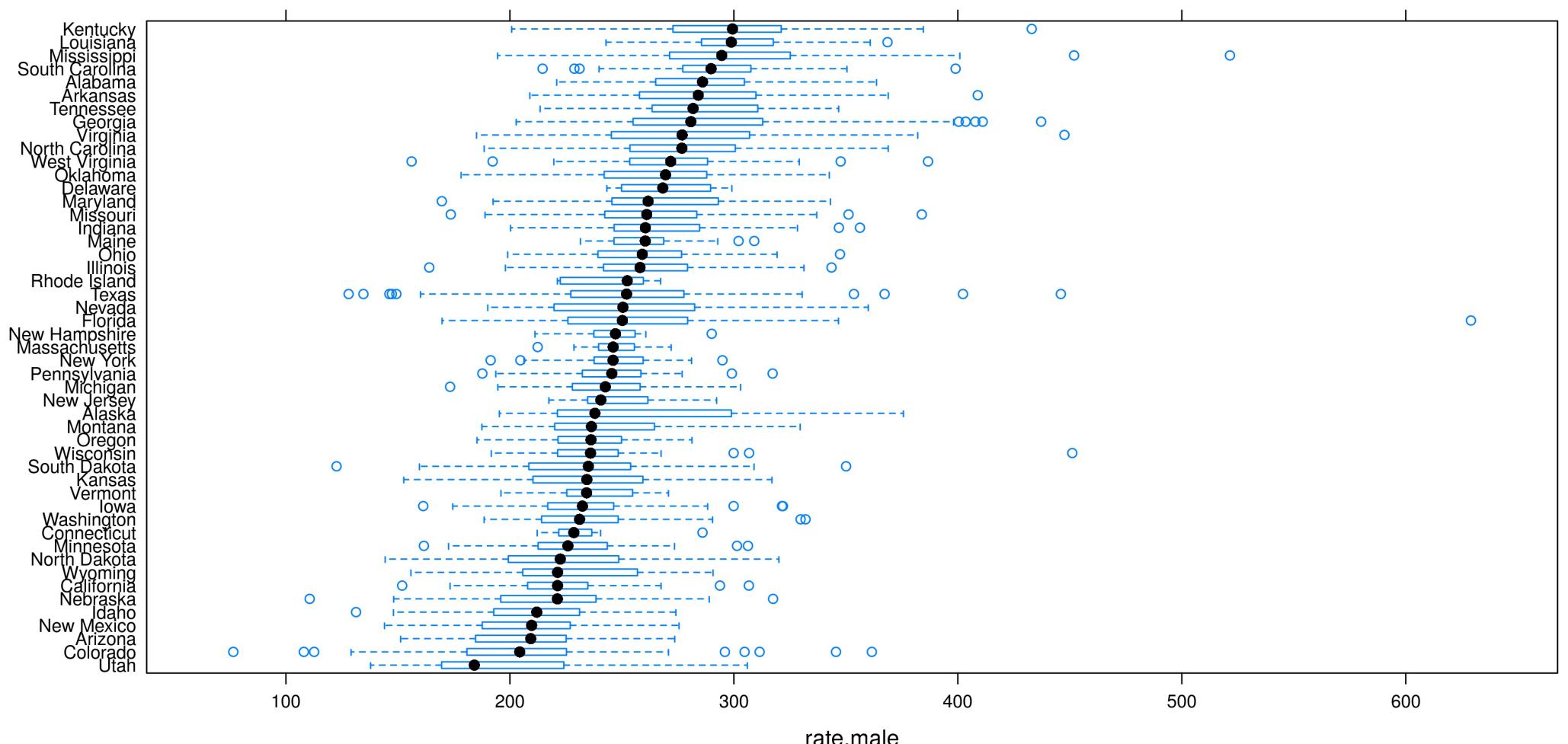
Comparative box and whisker plots

```
> bwplot(state ~ rate.male, data = USCancerRates)
```



Reordering factor levels

```
> library(dplyr)  
> USCancerRates <- mutate(USCancerRates,  
+   state.ordered = reorder(state, rate.male, median, na.rm = TRUE))  
> bwplot(state.ordered ~ rate.male, data = USCancerRates)
```





DATA VISUALIZATION IN R WITH LATTICE

Let's practice!