



DATA VISUALIZATION IN R

# **The world of data visualization**

# Graphical tools help us understand a dataset

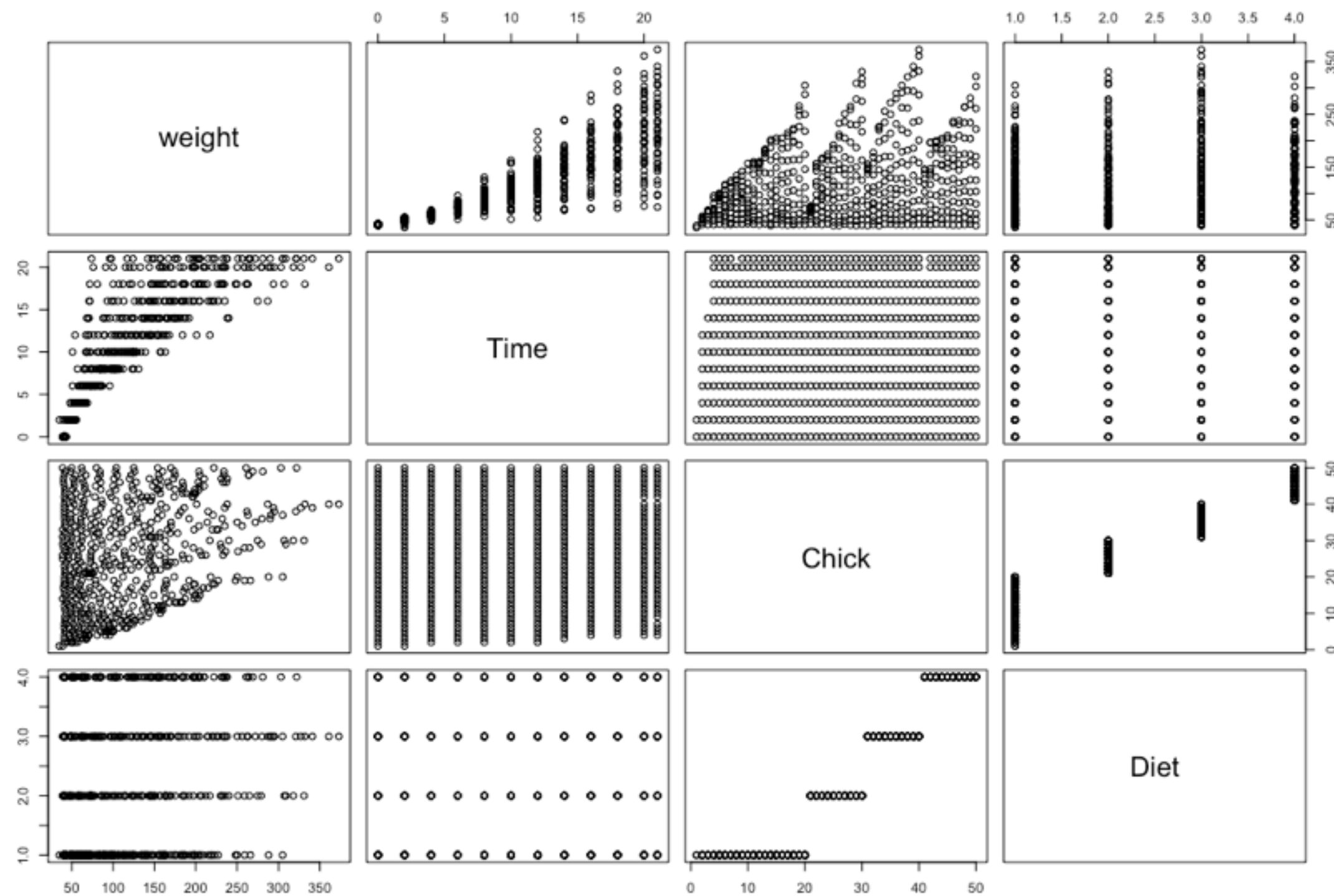
- Our ability to interpret patterns is a key strength
- Two basic types of data visualizations:
  - *Exploratory visualizations* help us understand the data
  - *Explanatory visualizations* help us share our understanding with others
- R supports both types of visualizations

# Exploratory data visualizations

- Helps *you* see what is in your data
- Level of detail:
  - Keep as much detail as possible
  - Practical limit: how much can *you* see and interpret?

# Exploratory data visualizations

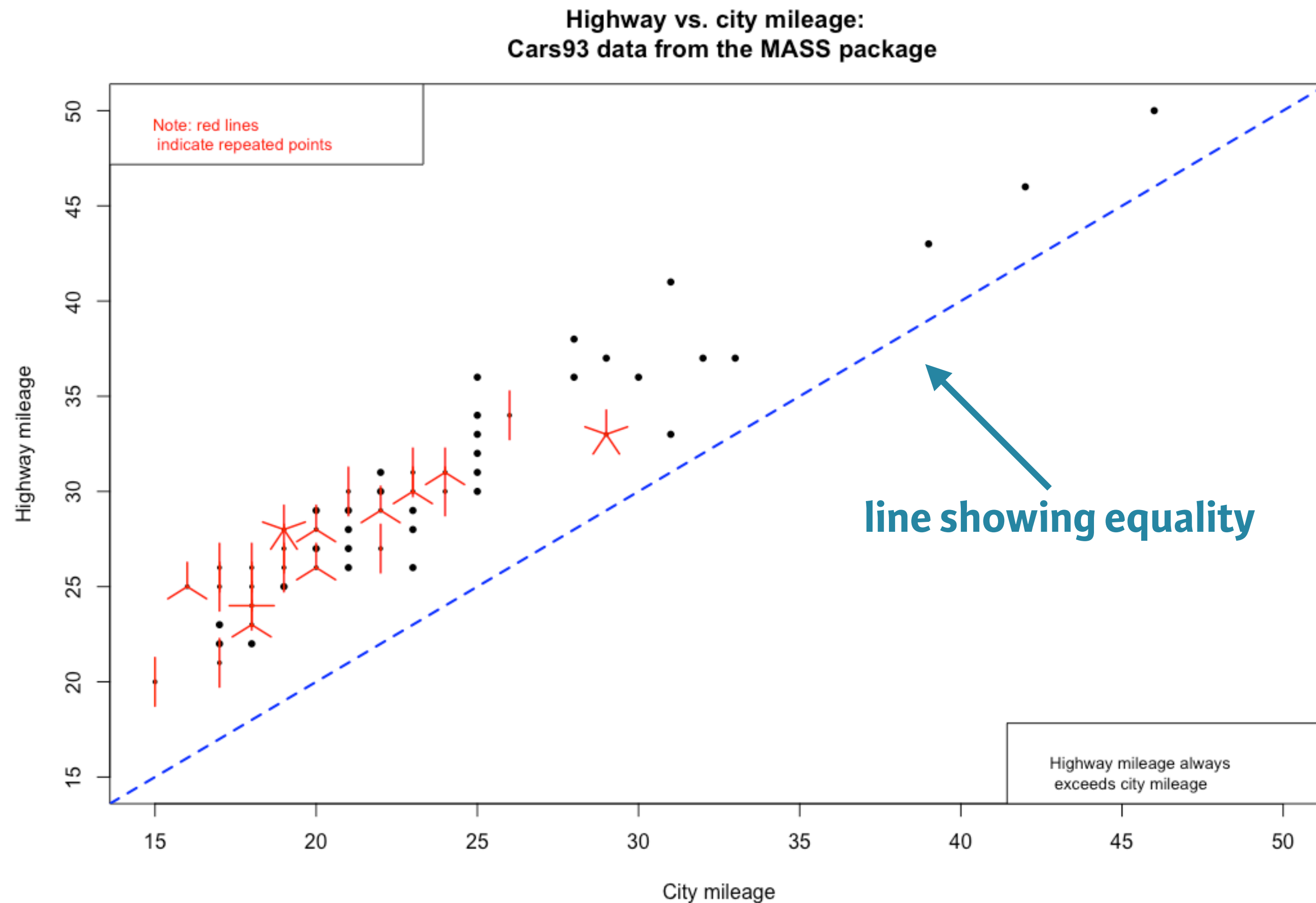
```
# Exploratory plot of ChickWeight data frame  
> plot(ChickWeight)
```



# Explanatory data visualizations

- Shows *others* what you've found in your data
- Requires editorial decisions:
  - Highlight the key features you want to emphasize
  - Eliminate extraneous details

# Explanatory data visualizations



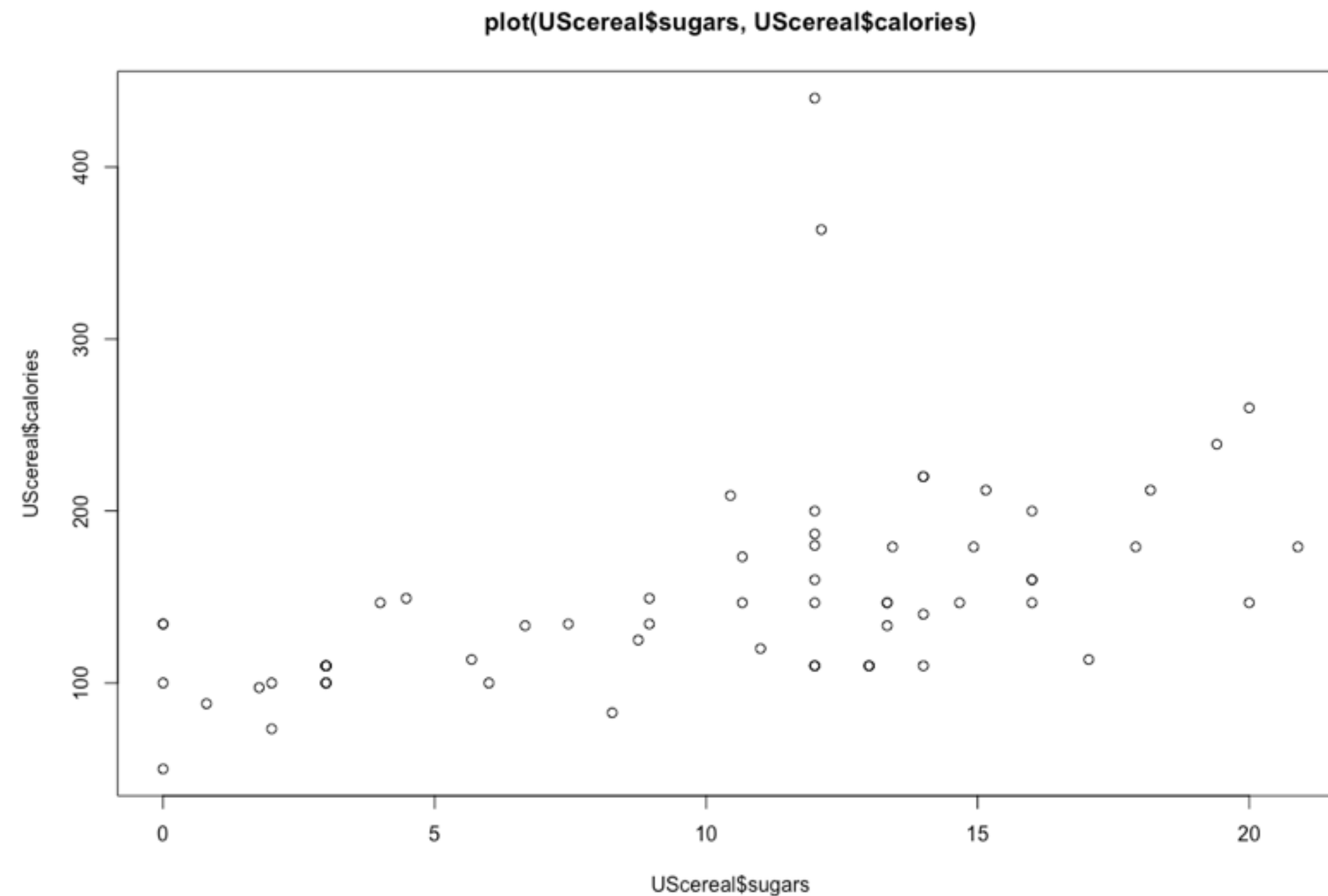
# Four graphics systems in R

- *Base graphics*: Easiest to learn and focus of this course
- *Grid graphics*: powerful set of modules for building other tools
- *Lattice graphics*: general purpose system based on grid graphics
- *ggplot2*: “the grammar of graphics”



# Base graphics example

```
> library(MASS)
> plot(UScereal$sugars, UScereal$Calories)
> title("plot(UScereal$sugars, UScereal$calories)")
```

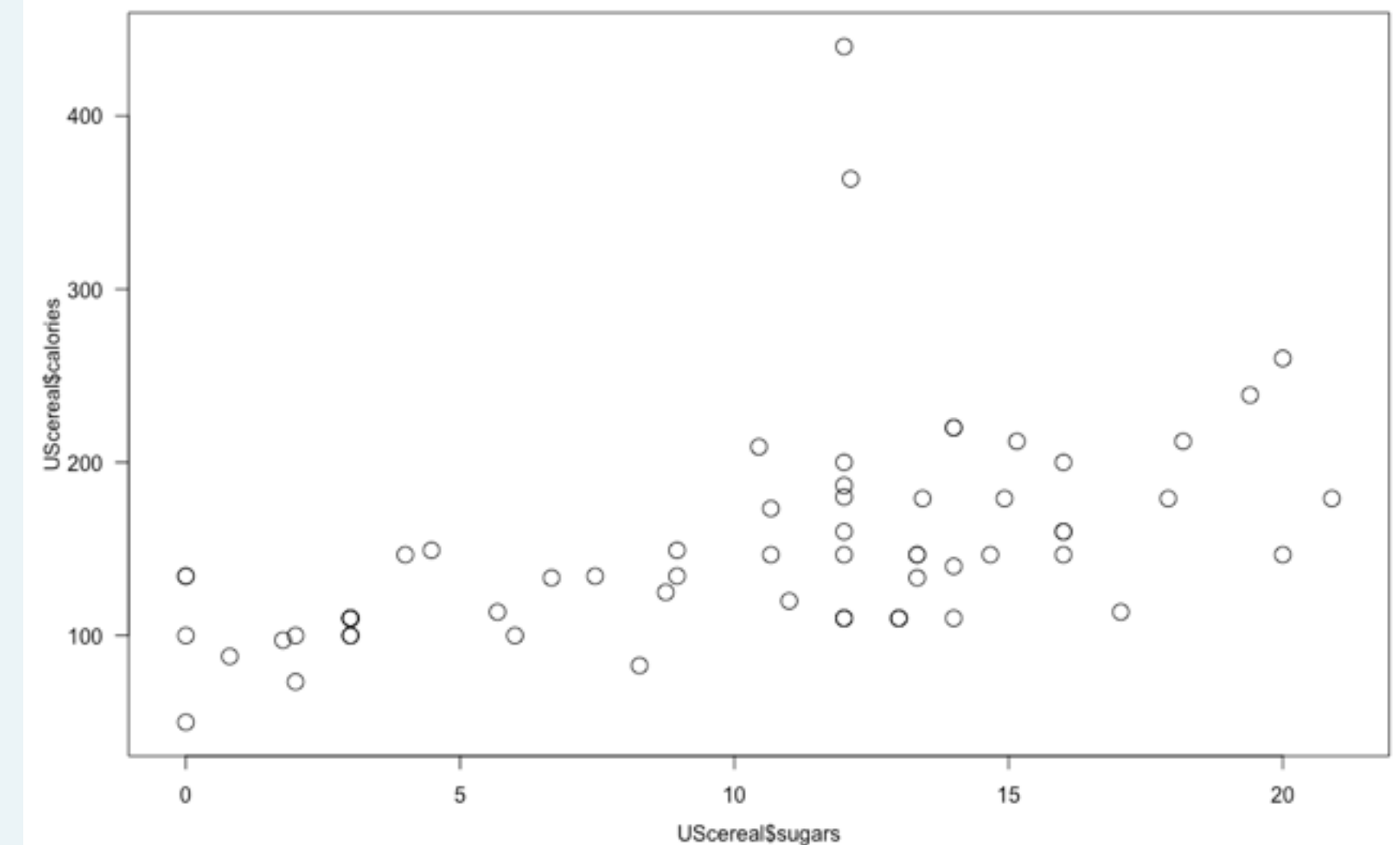




# Near equivalent in grid graphics

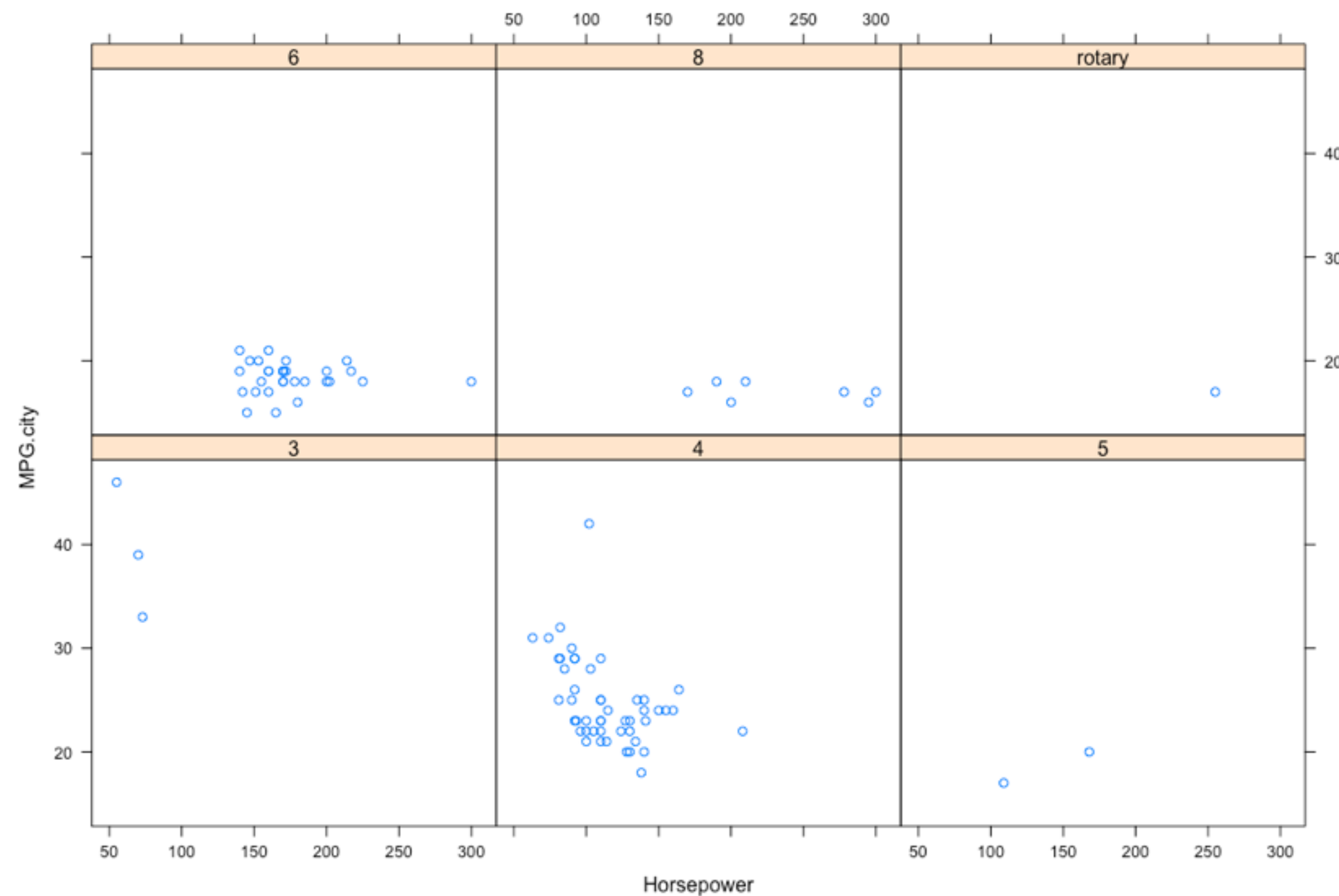
```
> # Get the data and load the grid package
> library(MASS)
> x <- UScereal$sugars
> y <- UScereal$calories
> library(grid)

> # This is the grid code required to generate the plot
> pushViewport(plotViewport())
> pushViewport(dataViewport(x, y))
> grid.rect()
> grid.xaxis()
> grid.yaxis()
> grid.points(x, y)
> grid.text("UScereal$calories", x = unit(-3, "lines"), rot = 90)
> grid.text("UScereal$sugars", y = unit(-3, "lines"), rot = 0)
> popViewport(2)
```



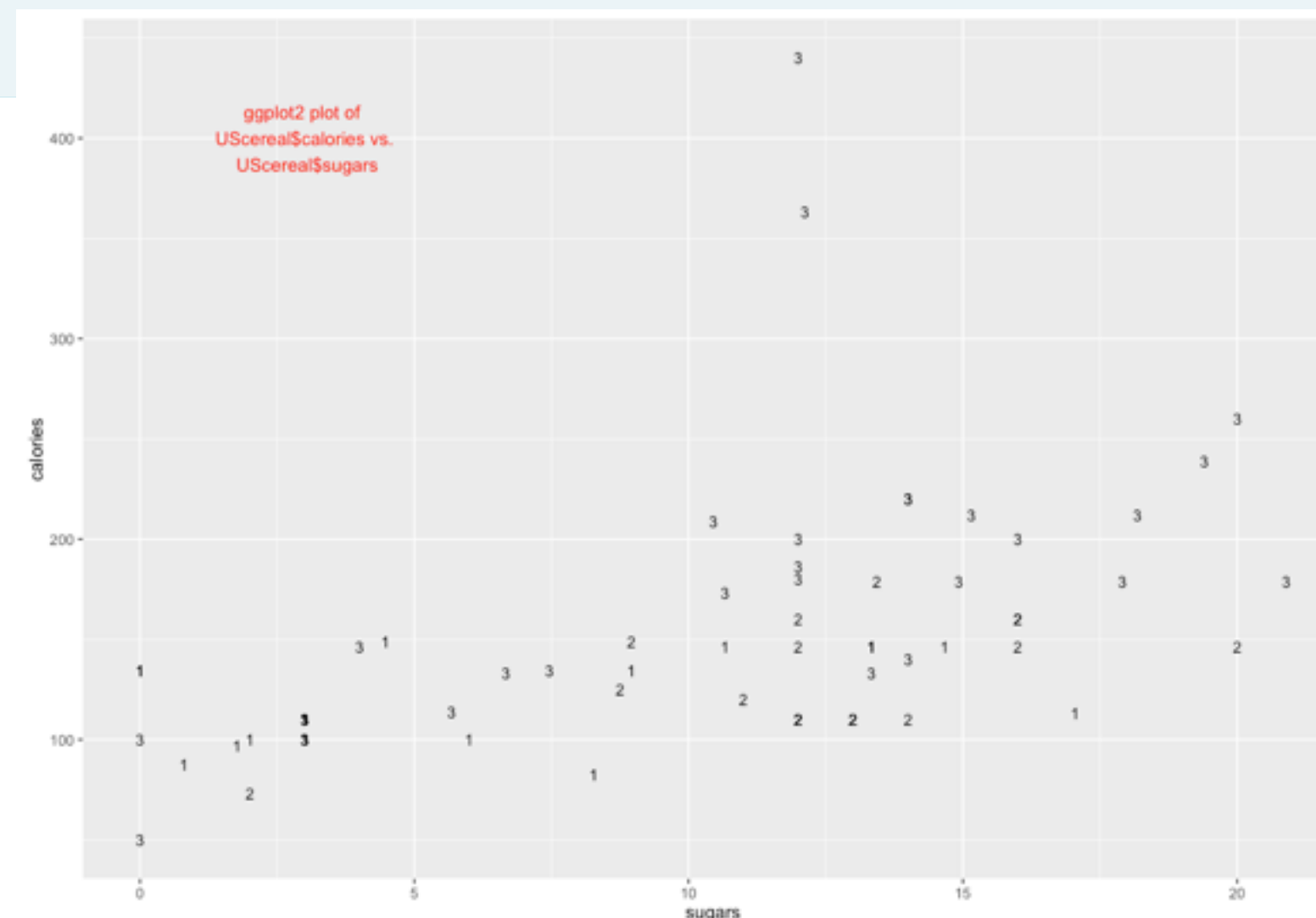
# Conditional scatterplot example from lattice graphics

```
> library(MASS)
> library(lattice)
> xyplot(MPG.city ~ Horsepower | Cylinders, data = Cars93)
```



# Example ggplot2 plot

```
> library(MASS)
> library(ggplot2)
> title <-
  "ggplot2 plot of \n UScereal$calories vs. \n UScereal$sugars"
> basePlot <- ggplot(UScereal, aes(x = sugars, y = calories))
> basePlot +
  geom_point(shape = as.character(UScereal$shelf), size = 3) +
  annotate("text", label = title, x = 3, y = 400,
          colour = "red")
```





DATA VISUALIZATION IN R

**Let's practice!**