Data visualization in R
Basic graphics

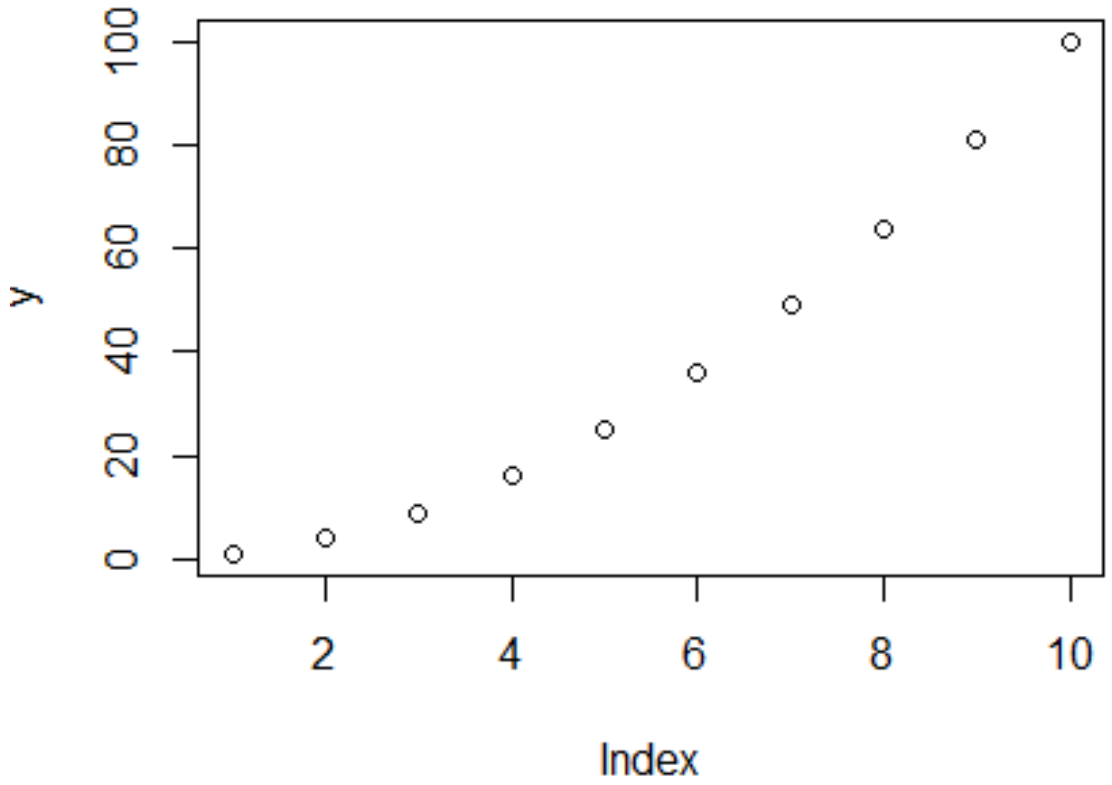
In this lecture

- Basic graphics
  - Scatter
  - Line
  - Bar
- Need for sophisticated graphics

Scatter plot

R – code:

```
X = 1:10
Y= X^2
plot (Y)
```



Scatter plot

dataset 'mtcars':

The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

Usage

```
mtcars
```

Format

A data frame with 32 observations on 11 variables.

[, 1]	mpg	Miles/(US) gallon
[, 2]	cyl	Number of cylinders
[, 3]	disp	Displacement (cu.in.)
[, 4]	hp	Gross horsepower
[, 5]	drat	Rear axle ratio
[, 6]	wt	Weight (1000 lbs)
[, 7]	qsec	1/4 mile time
[, 8]	vs	V/S
[, 9]	am	Transmission (0 = automatic, 1 = manual)
[,10]	gear	Number of forward gears
[,11]	carb	Number of carburetors

Source

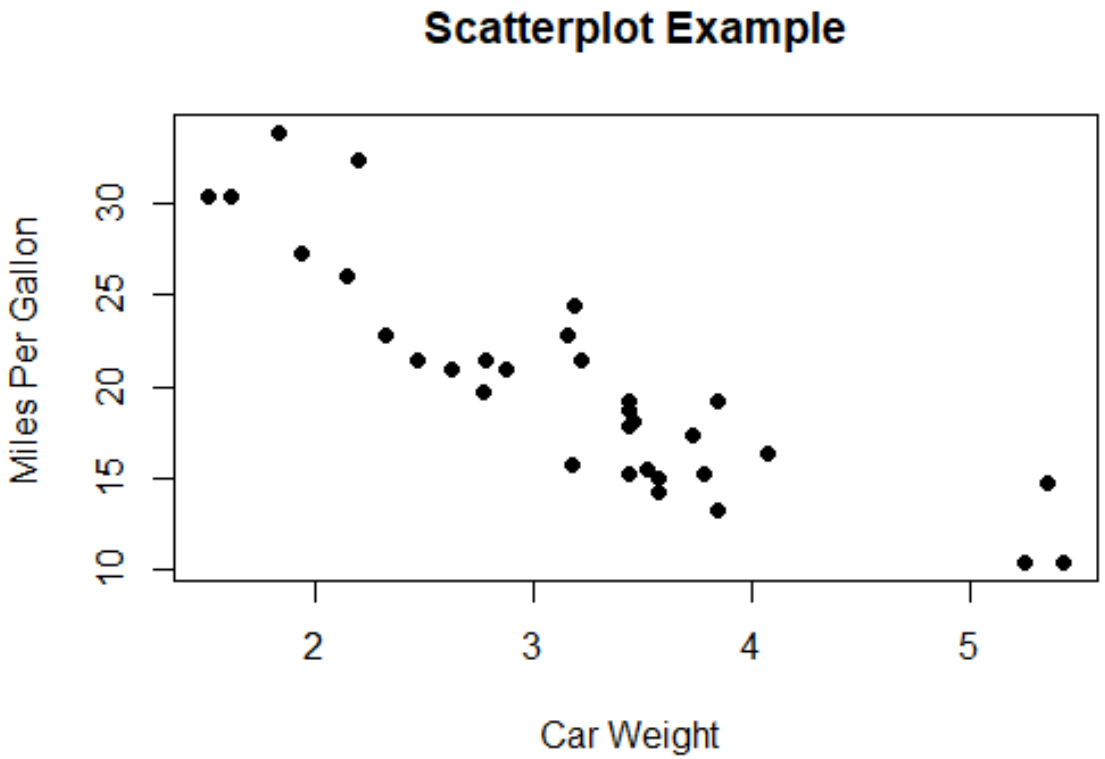
Henderson and Velleman (1981), Building multiple regression models interactively. *Biometrics*, **37**, 391–411.

Scatter plot

R – code :

```
plot( mtcars$wt, mtcars$mpg ,
main="Scatterplot Example",
   xlab="Car Weight ", ylab="Miles Per Gallon ", pch=19)
```

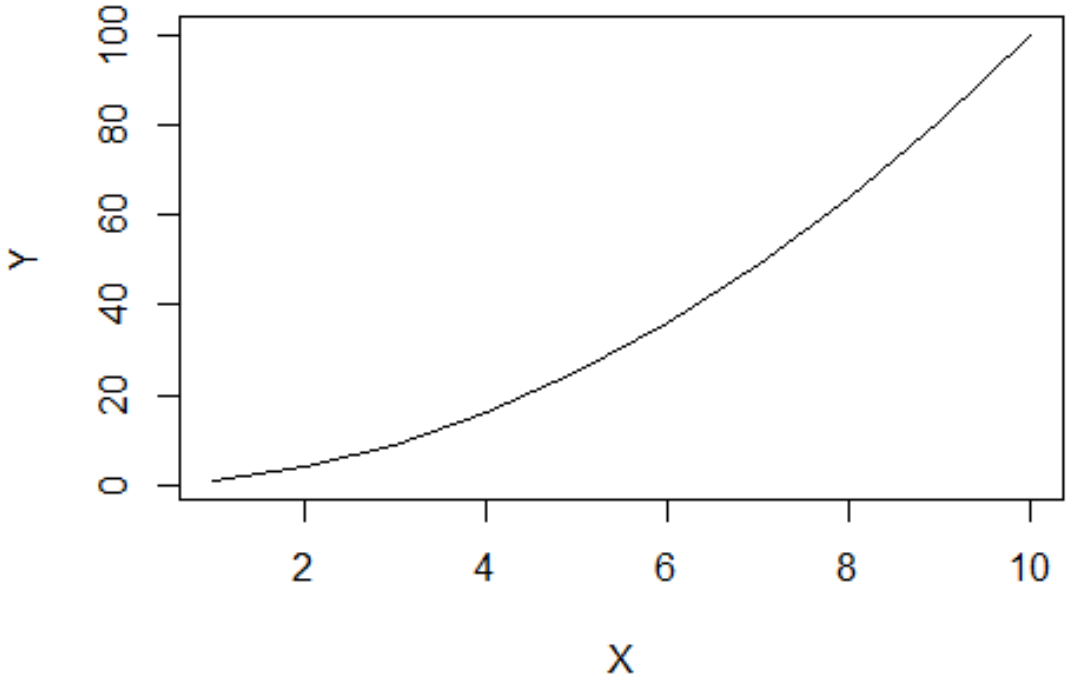
Corresponds to different shapes for points, for more such options check 'graphics parameters' in help



Line plot

R – code :

```
X = 1:10
Y= X^2
plot(X,Y,type = 'l')
```



Bar plot

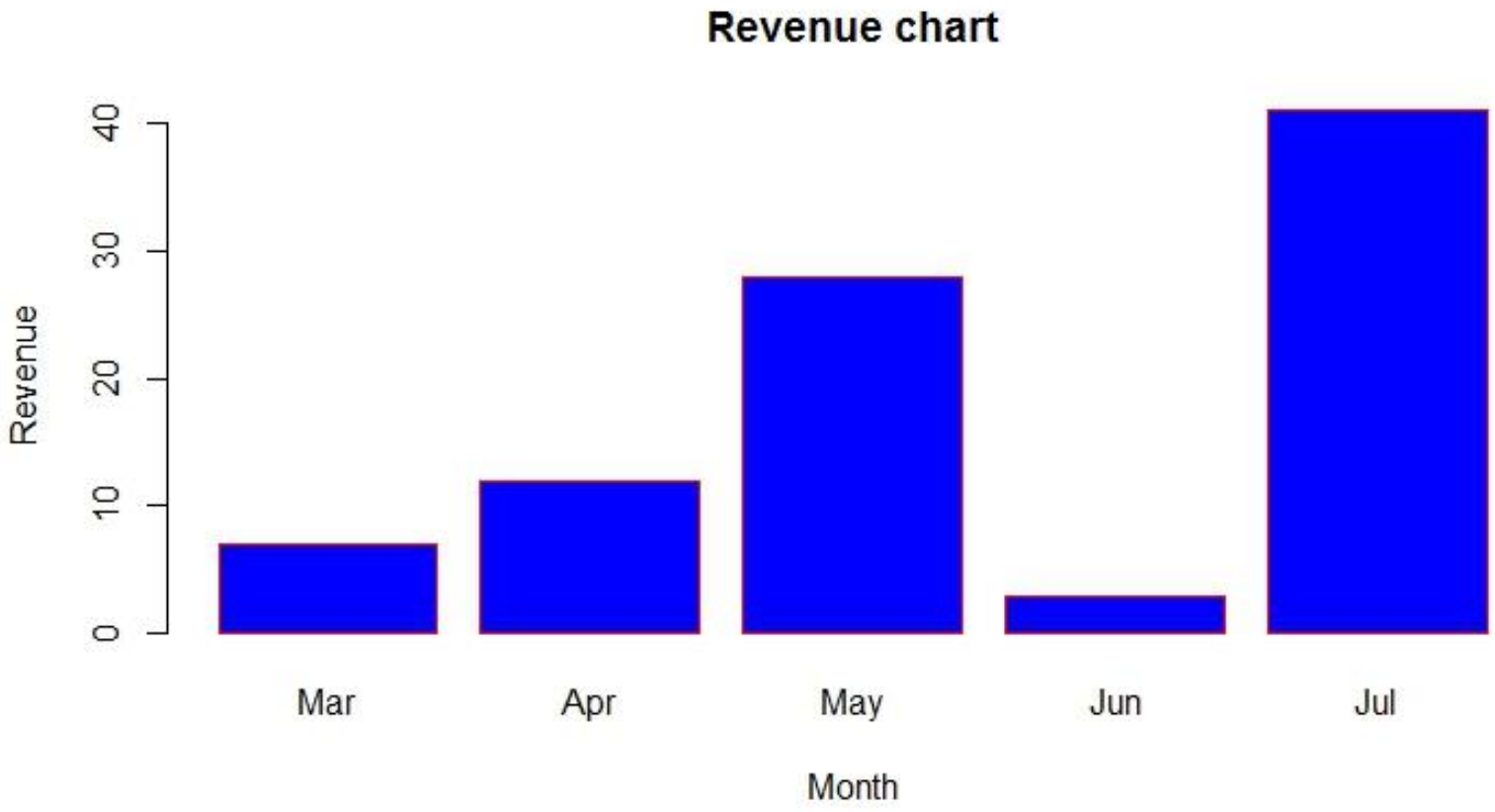
Syntax:

barplot(H, names.arg, xlab, ylab, main, names.arg, col)

R – code :

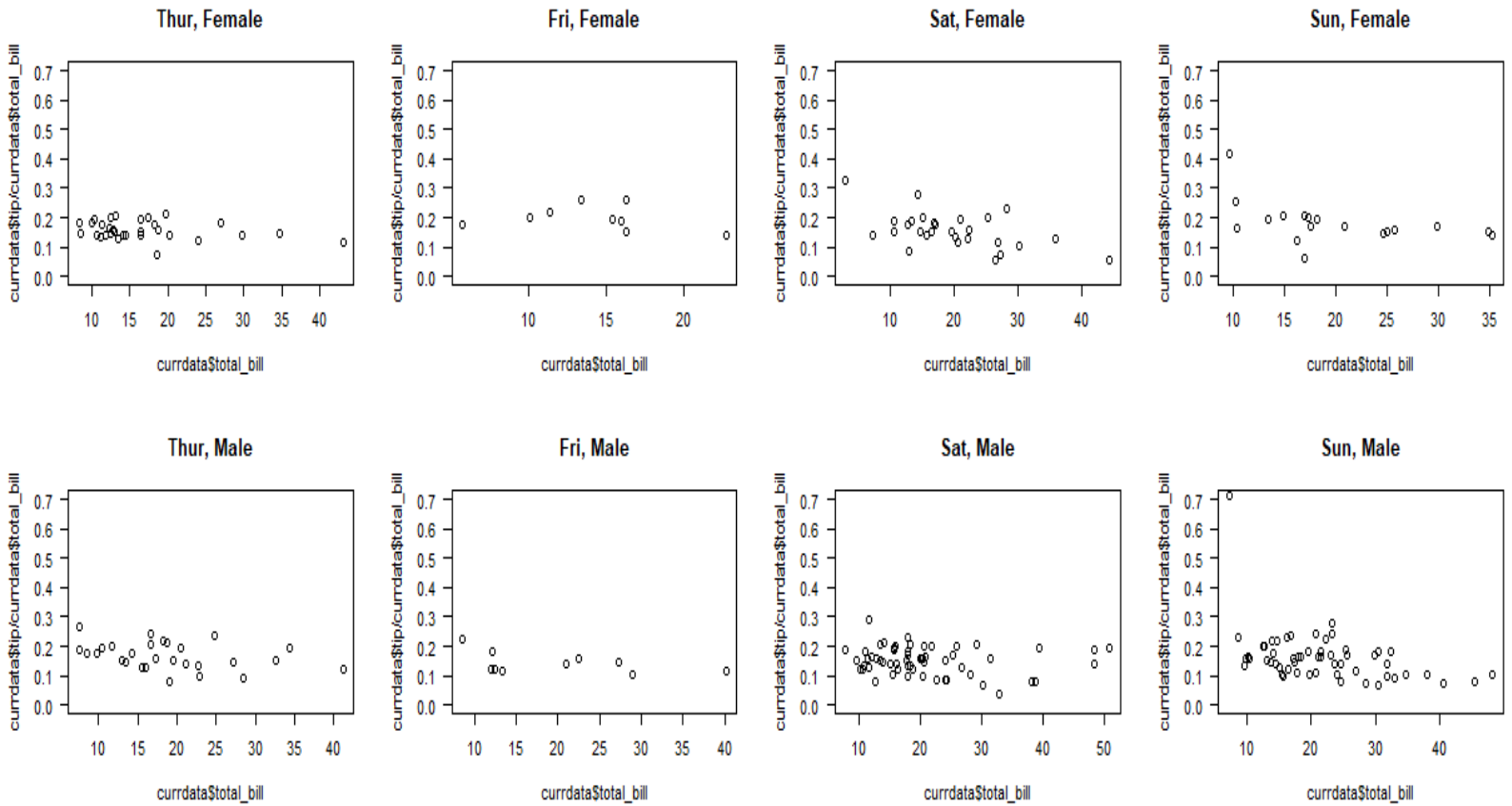
```
H <- c(7,12,28,3,41)
M <- c("Mar","Apr","May","Jun","Jul")
barplot(H,names.arg = M, xlab = "Month", ylab = "Revenue",
col = "blue", main = "Revenue chart",border = "red")
```

Bar plot



Need for sophisticated graphics

Lets us say there is a need for you to show multiple plots in a single figure such as the following:



Challenges

The exact figure as per the previous slide can be reproduced with the following code:

```
par(mfrow=c(2,4))
days <- c("Thur", "Fri", "Sat", "Sun")
sexes <- unique(tips$sex)
for (i in 1:length(sexes)) {
  for (j in 1:length(days)) {
    currdata <- tips[tips$day == days[j] & tips$sex == sexes[i],]
    plot(currdata$total_bill, currdata$tip/currdata$total_bill,
        main=paste(days[j], sexes[i], sep=", "), ylim=c(0,0.7), las=1)
  }
}
```

Challenges

But the code requires work such as :

- ***Knowing when to introduce a for loop***
- ***Which columns of the data.frame to select***
- ***The positioning of each graph in the grid etc***
- ***Less pleasing visuals***

Summary

1) Scatter plots
2) Line plots
3) Bar plots
4) Challenges and disadvantages of basic graphics