Econometrics 1 *Applied Econometrics with R*

Lecture 3: Basics of R — Part 2

黄嘉平

中国经济特区研究中心 讲师

办公室：文科楼2613

E-mail: huangjp@szu.edu.cn

Tel: (0755) 2695 0548

Website: https://huangjp.com

Review

1. Basic operators and functions:
   `+, -, *, /, ^, exp(), log(), mean(), var(), sum()`, etc.

2. Vectors and matrices
   - `c()`, `length()`, `matrix()`, `seq()`, etc.
   - `()` and `[]`.
   - `%*%`, `solve()`, `t()`, etc.

Why `(-8)^(1/3)` leads to `NaN`

- `> (-8)^(1/3)↩`
  `[1] NaN`

- The definition of *n*th root: $r = x^{\frac{1}{n}} \Leftrightarrow r^n = x$
  Therefore, every number *x* has *n* (*n* is a positive integer) **real** or **complex** *n*th roots.

- The three cube roots of -8 are: $-2,\ 1+\sqrt{3}i,\ 1-\sqrt{3}i$

  principle root

- `> (-8+0i)^(1/3)↩`
  `[1] 1+1.732051i`

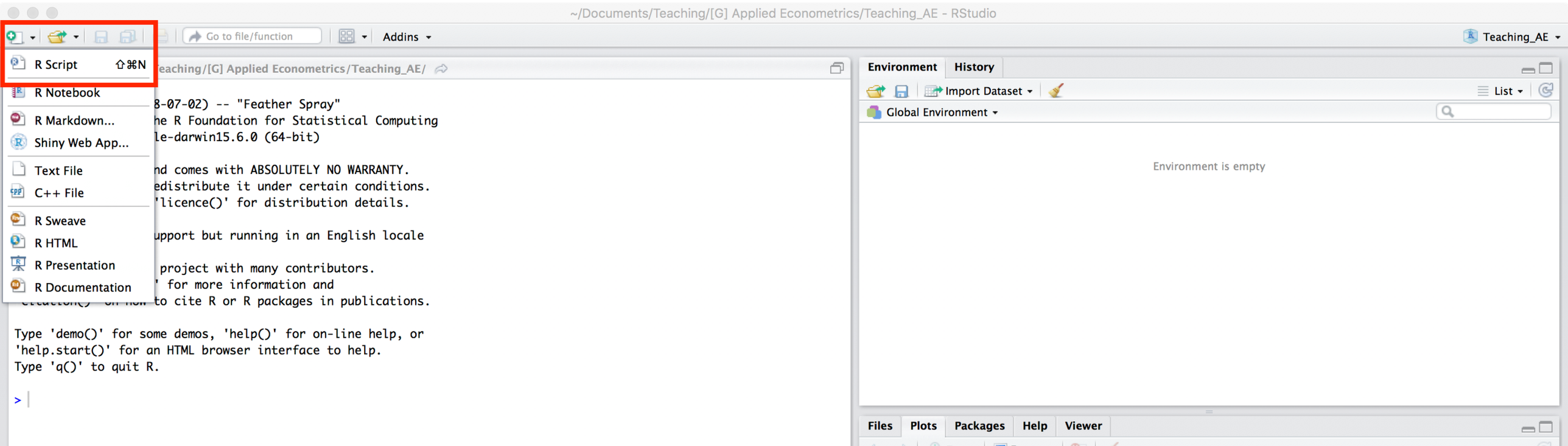
Save your code in a script file

Save your code in a script file

- One advantage of R is that it is interactive
- You can save your entire workspace
- It is highly recommended that you save your script

An R script is a text file with extension .R

Use `source()` to run a script file.
You can also select some lines and click “Run”.


~/Documents/Teaching/[G] Applied Econometrics/Teaching_AE - RStudio
Go to file/function
Addins
Teaching_AE
R Script ⇧⌘N
R Notebook
R Markdown...
Shiny Web App...
Text File
C++ File
R Sweave
R HTML
R Presentation
R Documentation
eaching/[G] Applied Econometrics/Teaching_AE/
8-07-02) -- "Feather Spray"
he R Foundation for Statistical Computing
le-darwin15.6.0 (64-bit)
nd comes with ABSOLUTELY NO WARRANTY.
edistribute it under certain conditions.
'licence()' for distribution details.
upport but running in an English locale
project with many contributors.
' for more information and
to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Zoom
Export

~/Documents/Teaching/[G] Applied Econometrics/Teaching_AE - RStudio

Go to file/function | Addins | Teaching_AE

Untitled1 | Source on Save | Run | Source

1

编辑窗口

1:1 (Top Level) R Script

Environment | History | Import Dataset | List | Global Environment

Environment is empty

Files | Plots | Packages | Help | Viewer | Zoom | Export

Console ~/Documents/Teaching/[G] Applied Econometrics/Teaching_AE/

```
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

Practice

- Matrix multiplication is not communicative, i.e.,

$$AB \neq BA$$

  Let us check this through an example.

- Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 0 & -1 \\ 6 & 7 \end{bmatrix}$,

  create a new script file, calculate AB and BA, and save your code.

Practice

In your script file, you may write

```
A <- matrix(c(1, 3, 2, 4), 2, 2)↩
B <- matrix(c(0, -1, 6, 7), 2, 2, byrow = TRUE)↩
X <- A %*% B↩
Y <- B %*% A↩
```

In the console

```
> X↩
     [,1] [,2]
[1,]   12   13
[2,]   24   25
> Y↩
     [,1] [,2]
[1,]   -3   -4
[2,]   27   40
```

Data management in R

Data frame

- A typical way of storing data in R is using data frame

  ```
  > library("AER")↩
  ```
  if not installed, use `install.packages("AER")`

```
> data("Journals")↩
> Journals↩
> class(Journals)↩
> str(Journals)↩
> summary(Journals)↩
> names(Journals)↩
> Journals$citations↩
```

Create a data frame

- Create data frame with new data

```
> mydata <- data.frame(one = 1:10, two=
11:20, three = 21:30)↩
```

- Converting data to data.frame

```
> mymatrix <- matrix(1:30, ncol = 3)↩
> mydata2 <- as.data.frame(mymatrix)↩
> names(mydata2) <- c(“one”, “two”,
“three”)↩
```

Data import: from excel to R

- Get a sample data:

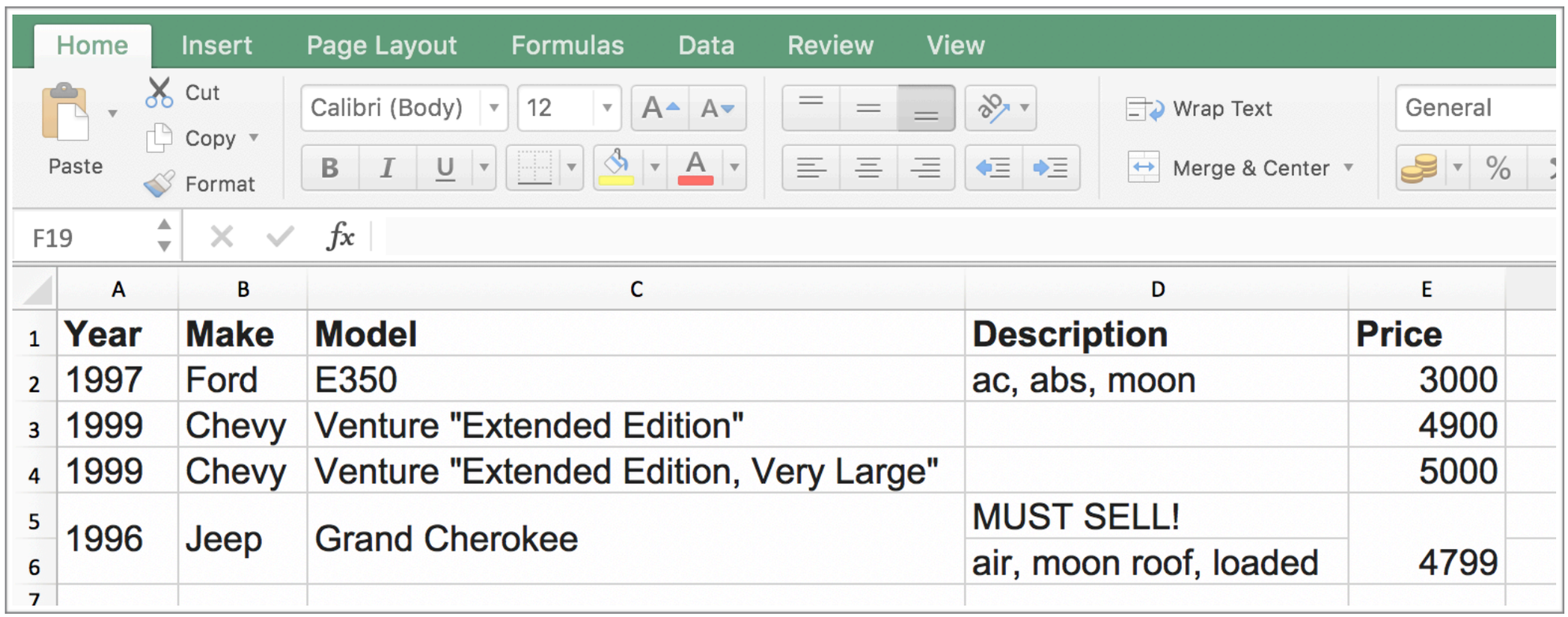
  广东统计年鉴2017
  `http://www.gdstats.gov.cn/tjsj/gdtjnj/`

  点击“查询”，将下载的rar文件解压后，打开
  `directory.html`

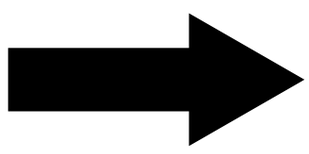
- Find your interested data, open and clean the excel data file and save it to a .csv file.

A typical .csv file

- https://en.wikipedia.org/wiki/Comma-separated_values



	A	B	C	D	E
1	**Year**	**Make**	**Model**	**Description**	**Price**
2	1997	Ford	E350	ac, abs, moon	3000
3	1999	Chevy	Venture "Extended Edition"		4900
4	1999	Chevy	Venture "Extended Edition, Very Large"		5000
5	1996	Jeep	Grand Cherokee	MUST SELL!	
6				air, moon roof, loaded	4799
7					



```
Year,Make,Model,Description,Price
1997,Ford,E350,"ac, abs, moon",3000
1999,Chevy,"Venture ""Extended Edition""",,4900
1999,Chevy,"Venture ""Extended Edition, Very Large""",,5000
1996,Jeep,Grand Cherokee,MUST SELL! ,4799
,,,"air, moon roof, loaded",
```

Data import

- Use `read.csv()` to import data from a .csv file.
- Suppose you have saved your data into a csv file named `mydatafile.csv`, then you can load it with

```
> mydata <- read.csv("mydatafile.csv")↩
```

- Try `class()`, `str()`, `summary()`, etc. with `mydata`.

Data export

- Use `write.csv()` to save a data frame or a matrix into a .csv file, e.g.,

```
> write.csv(mydata, file = "mdfile.csv")↩
```

- Data can also be saved as a binary file.

```
> save(mydata, file = “mydatafile.RData”)↩
```

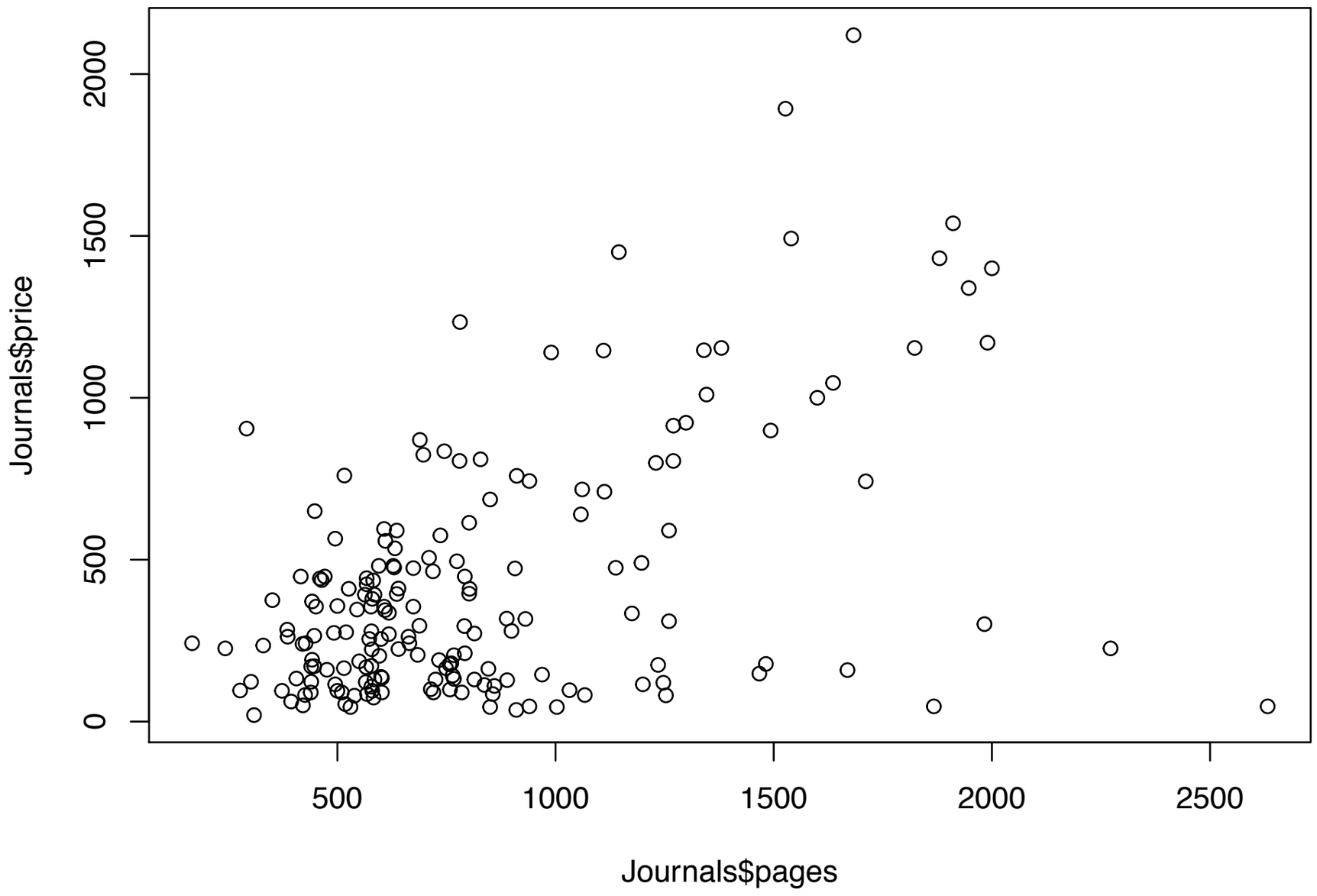
- Load .RData type of data

```
> load(“mydatafile.RData”)↩
```

Graphics

Basic plotting function: `plot()`

- Scatter plot is probably the most common graphical display in statistics.



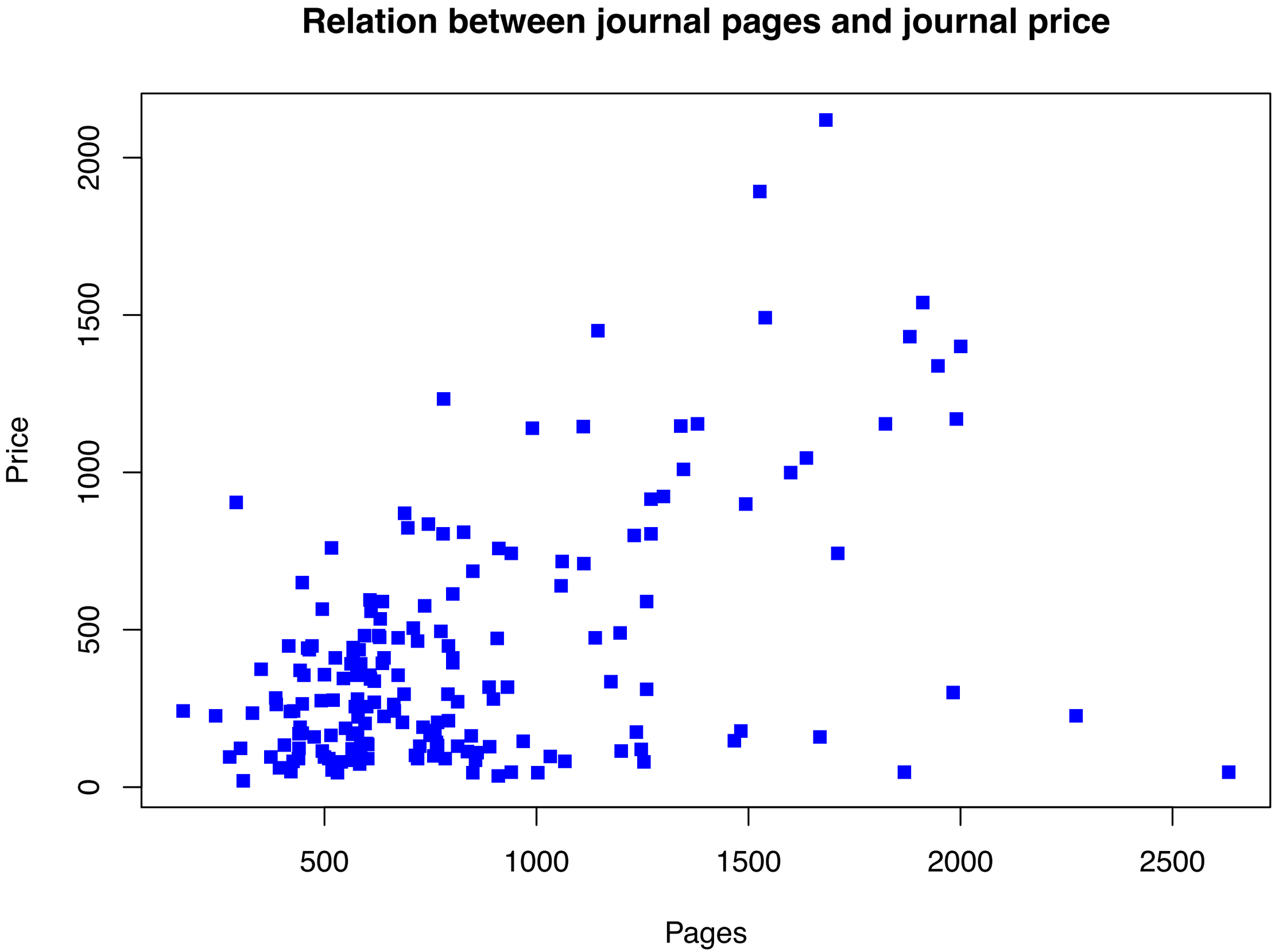
- Horizontal axis x, vertical axis y → `plot(x, y)`

```
> plot(Journals$pages,
Journals$price)↩
```

- Specify graphical parameters

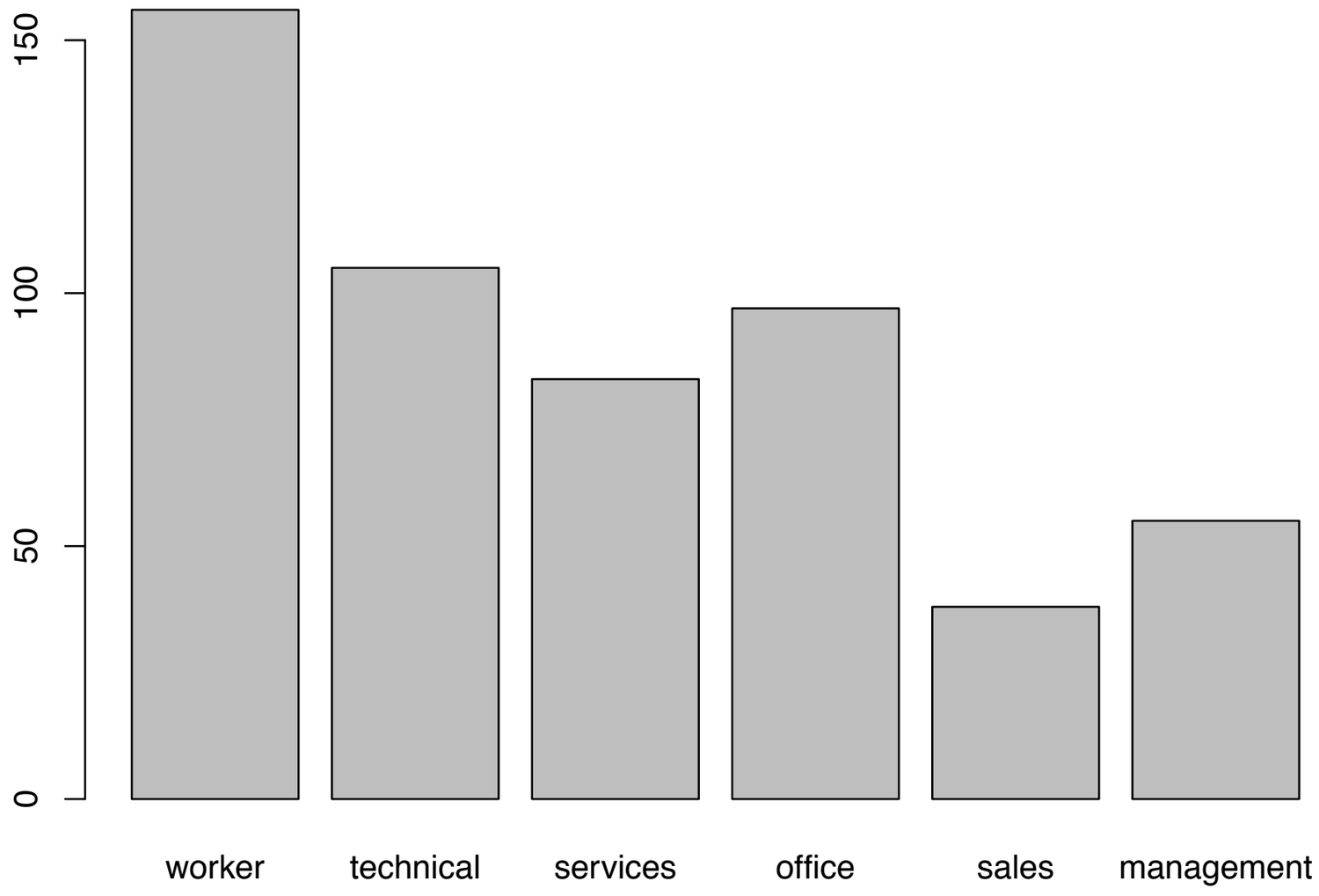
```
> plot(Journals$pages, Journals$price,
main = "Relation between journal pages
and journal price", xlab = "Pages",
ylab = "Price", col = "blue", pch =
15)↩
```

For further details, use `?plot`


Relation between journal pages and journal price
Price
0
500
1000
1500
2000
500
1000
1500
2000
2500
Pages

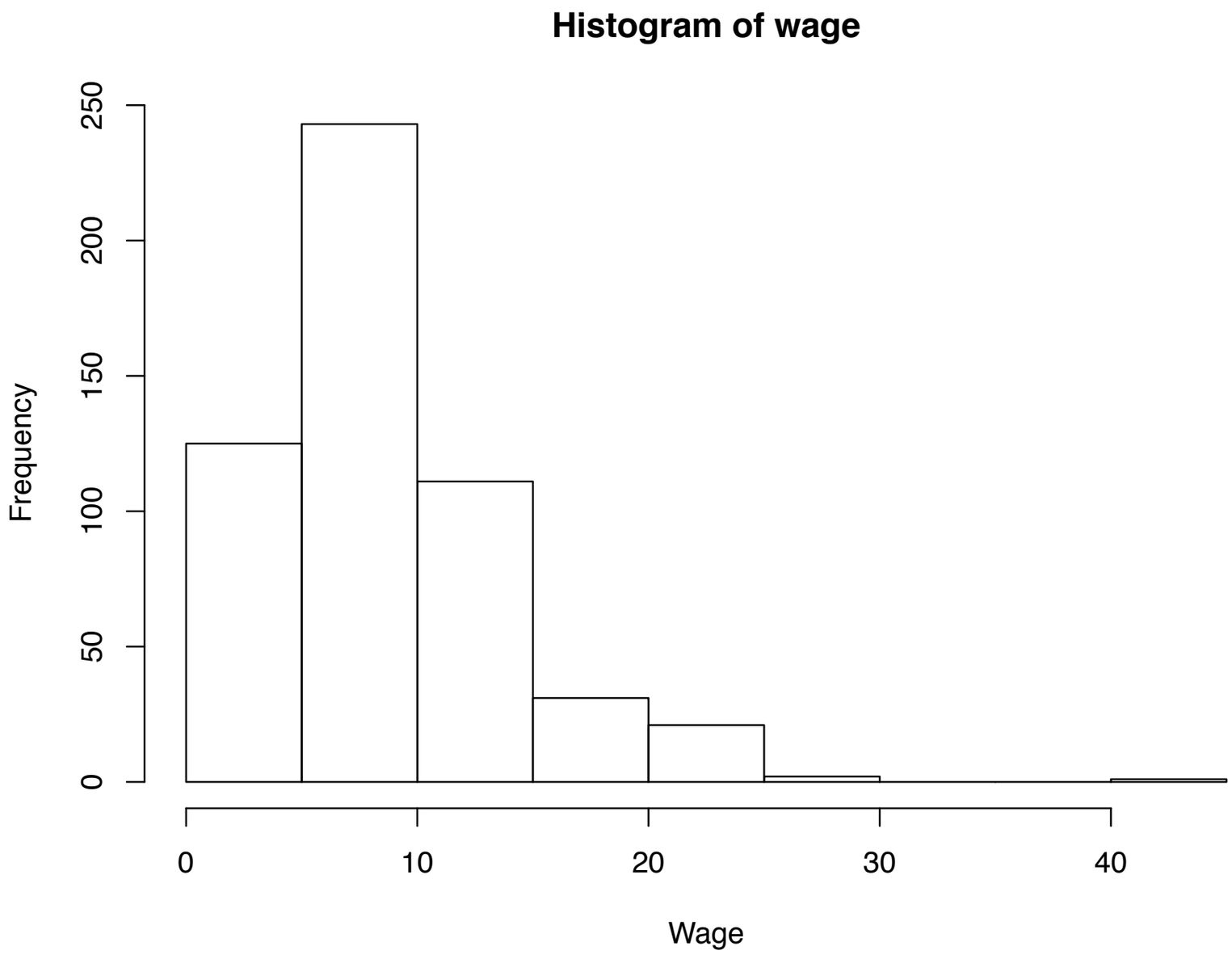
Bar graph

```
> data(CPS1985, package = "AER")↩
> barplot(summary(CPS1985$occupation))↩
```



Histogram

```
> hist(CPS1985$wage, main = "Histogram of
wage", xlab = "Wage")↩
```



Plot a function (1): a general way

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- Step 1: specify the range of your function

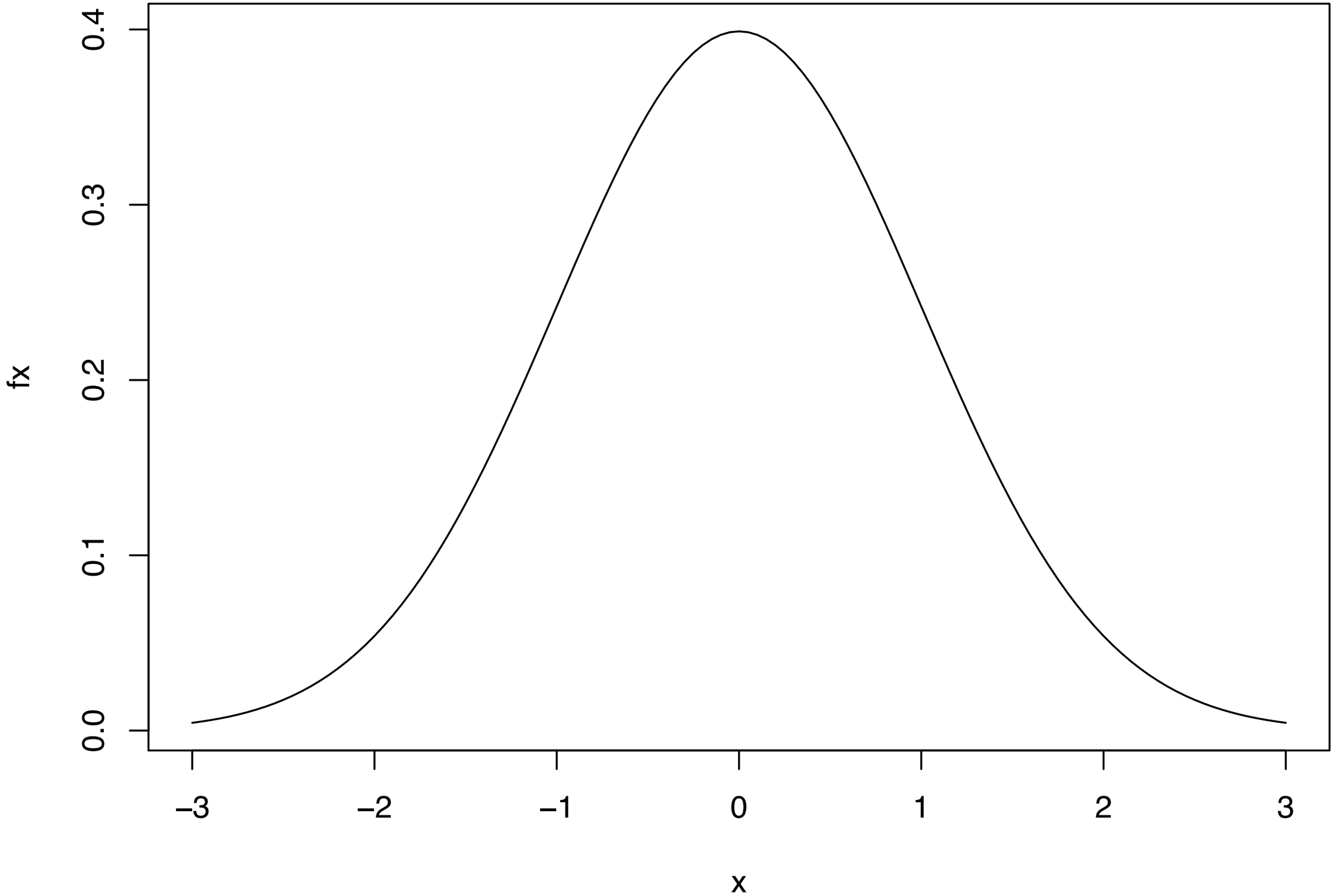
```
> x <- seq(-3, 3, 0.05)↩
```

- Step 2: calculate the values over the range

```
> fx <- exp(-x^2/2) / sqrt(2*pi)↩
```

- Step 3: plot

```
> plot(x, fx, type = "l")↩
```


fx
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
x

Plot a function (1): a faster way

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- We have plotted the density function of the standard normal distribution.

- This function can be calculated with the built-in command `dnorm()`

- Use `curve()` to draw the function

```
> curve(dnorm, from = -3, to = 3)↩
```

- Try it!

References

1. Kleiber, C. and Zeileis, A., *Applied Econometrics with R*, Springer, 2008.

2. Venables, W. N., Smith, D. M., and the R Core Team, *An Introduction to R*, Version 3.5.1, 2018-07-02.

https://cran.r-project.org/manuals.html