

# STAT UN1201 – Chapter 1

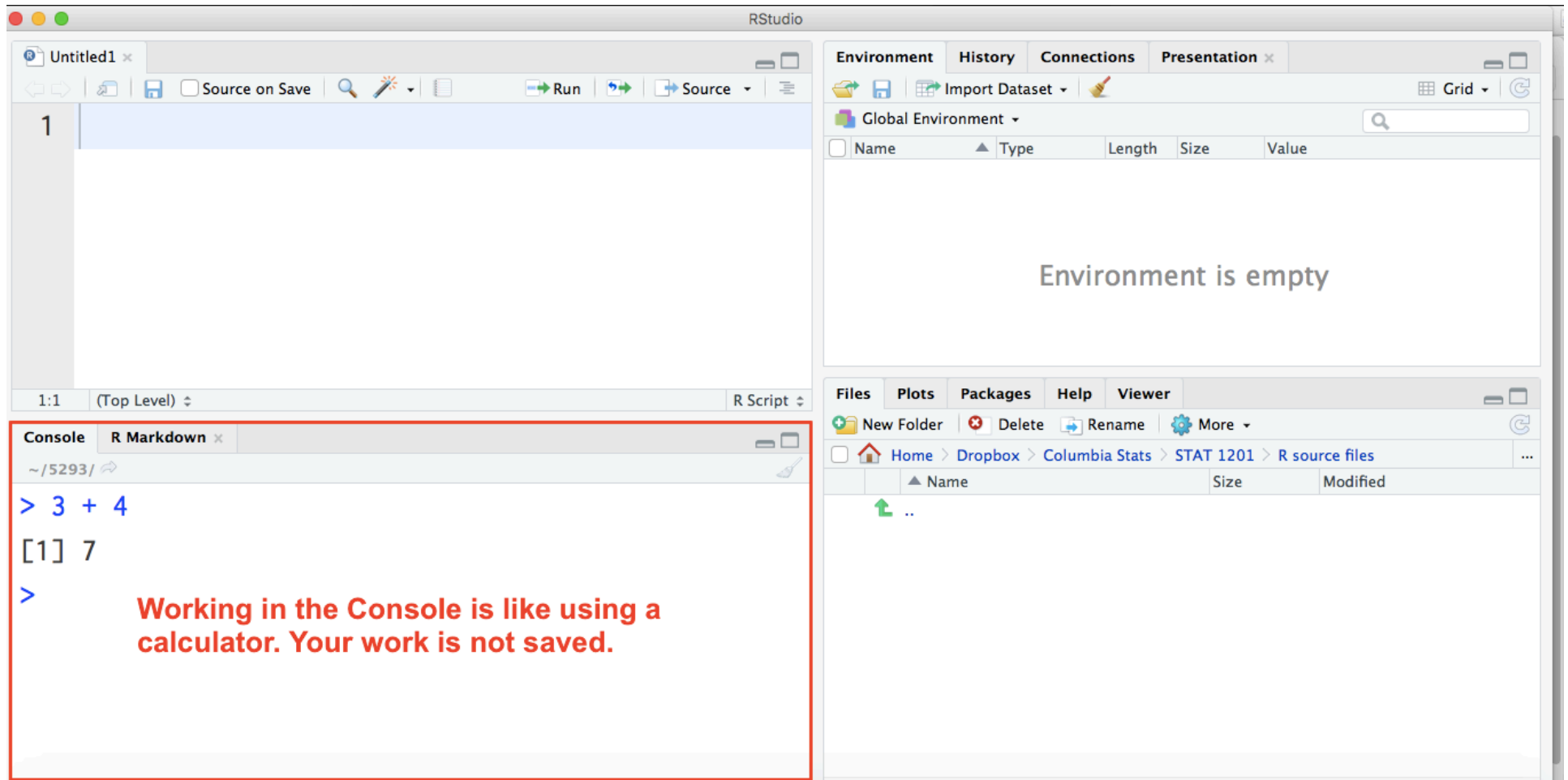
Prof. Joyce Robbins

# Waitlist

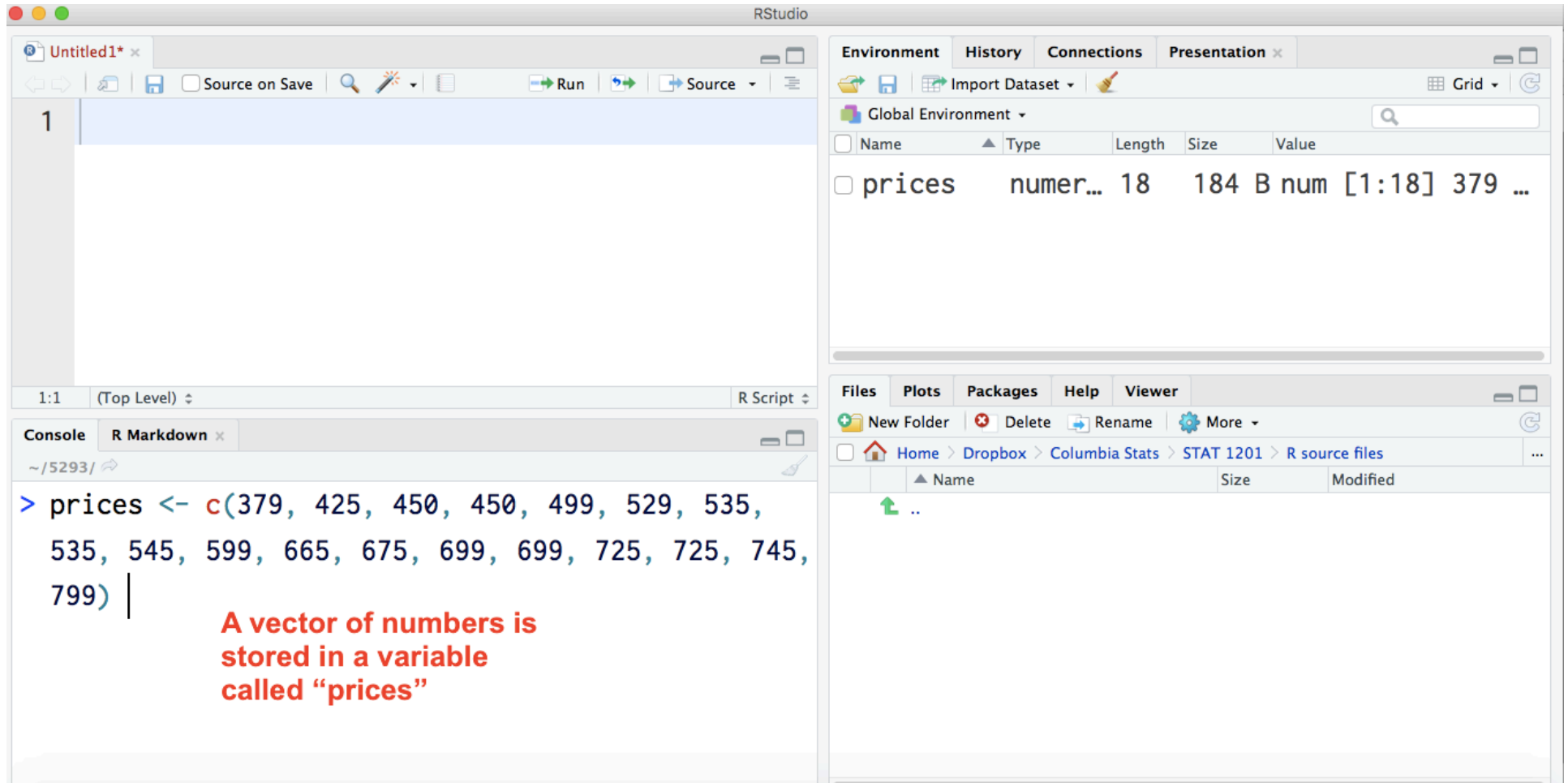
1. The waitlist moves in order as places open up.
2. Course materials are available here during change of program period: <http://github.com/jtr13/1201>
3. It is strongly advised to keep up with the material if you are trying to get in the class.

EVERYONE: Once you've made a decision not to take the class, please be considerate and drop it from your schedule.

# RStudio



# Working in the Console



The screenshot displays the RStudio environment. The top-left pane shows a script editor with a single line of R code: `prices <- c(379, 425, 450, 450, 499, 529, 535, 535, 545, 599, 665, 675, 699, 699, 725, 725, 745, 799)`. The bottom-left pane is the Console, showing the execution of this code. The right-hand side of the interface contains two panes. The top pane, titled 'Environment', shows the 'Global Environment' with a table listing the variable 'prices' as a numeric vector of length 18. The bottom pane, titled 'Files', shows the file explorer with the path 'Home > Dropbox > Columbia Stats > STAT 1201 > R source files'.

**Console Output:**

```
> prices <- c(379, 425, 450, 450, 499, 529, 535, 535, 545, 599, 665, 675, 699, 699, 725, 725, 745, 799)
```

**Environment Pane:**

Name	Type	Length	Size	Value
prices	numer...	18	184 B	num [1:18] 379 ...

**Files Pane:**

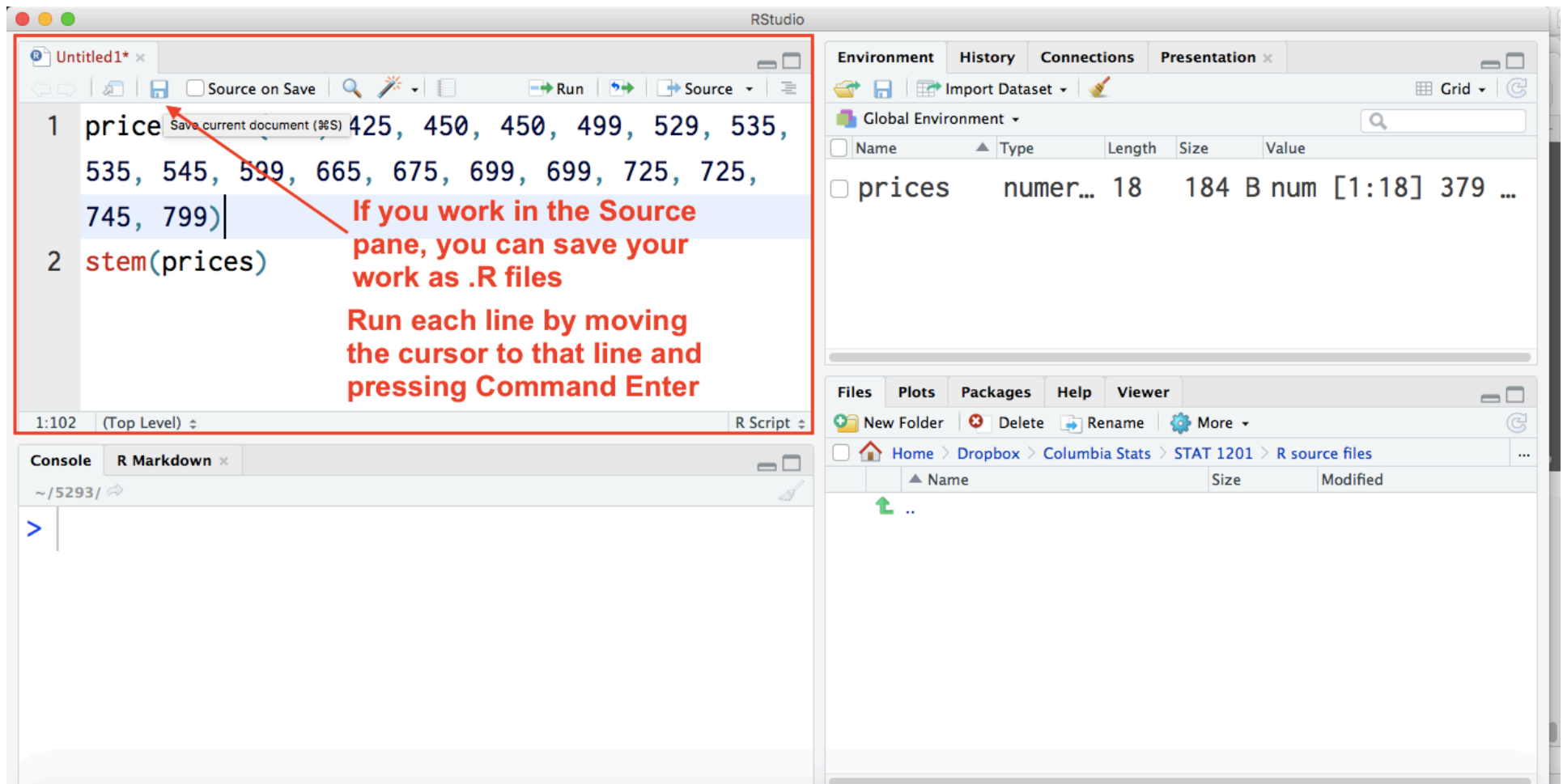
Home > Dropbox > Columbia Stats > STAT 1201 > R source files

..

**Annotation:**

A vector of numbers is stored in a variable called "prices"

# Saving your work as an R script



The screenshot shows the RStudio interface with the following components:

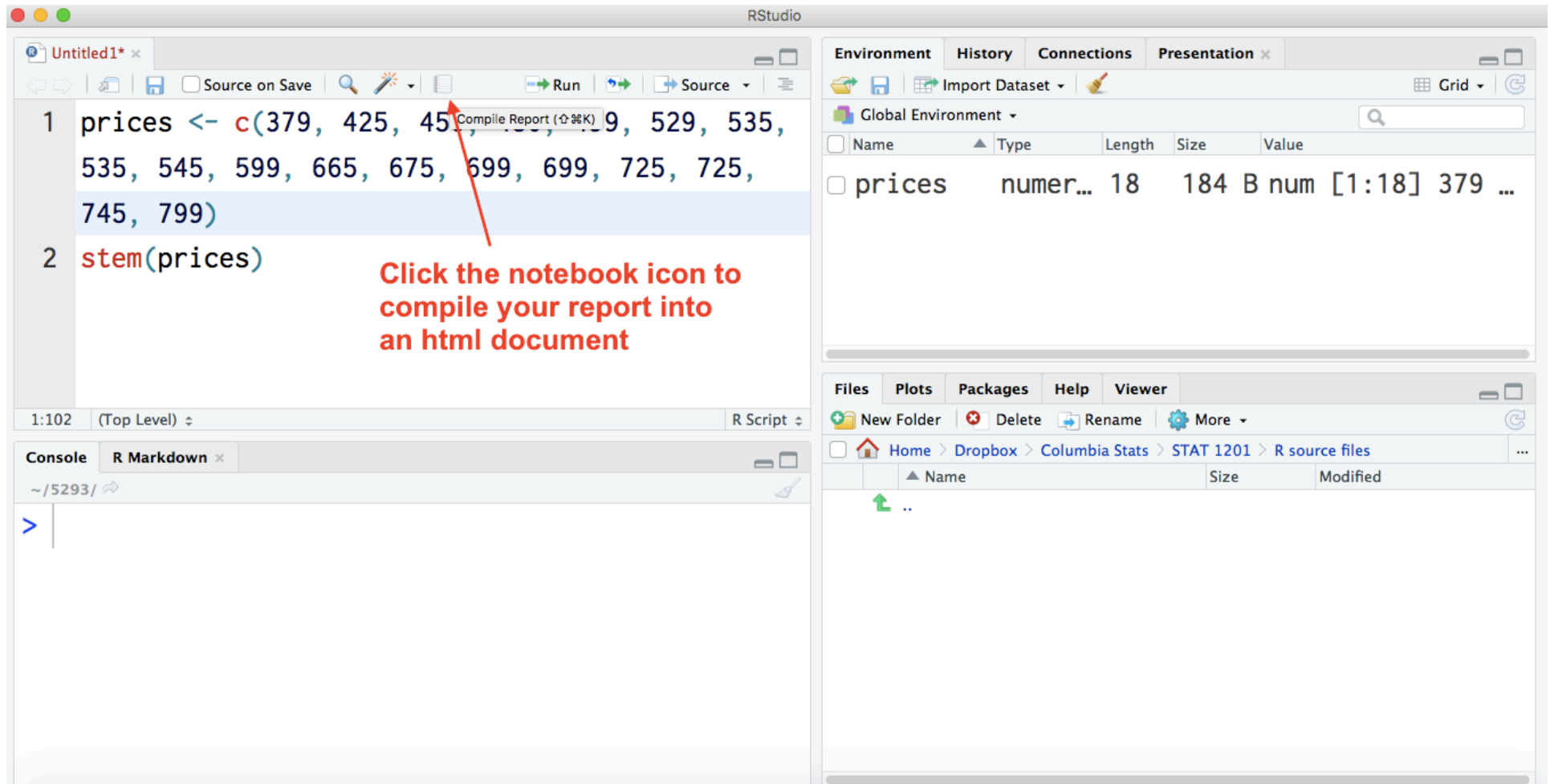
- Source Pane:** Contains two lines of R code:

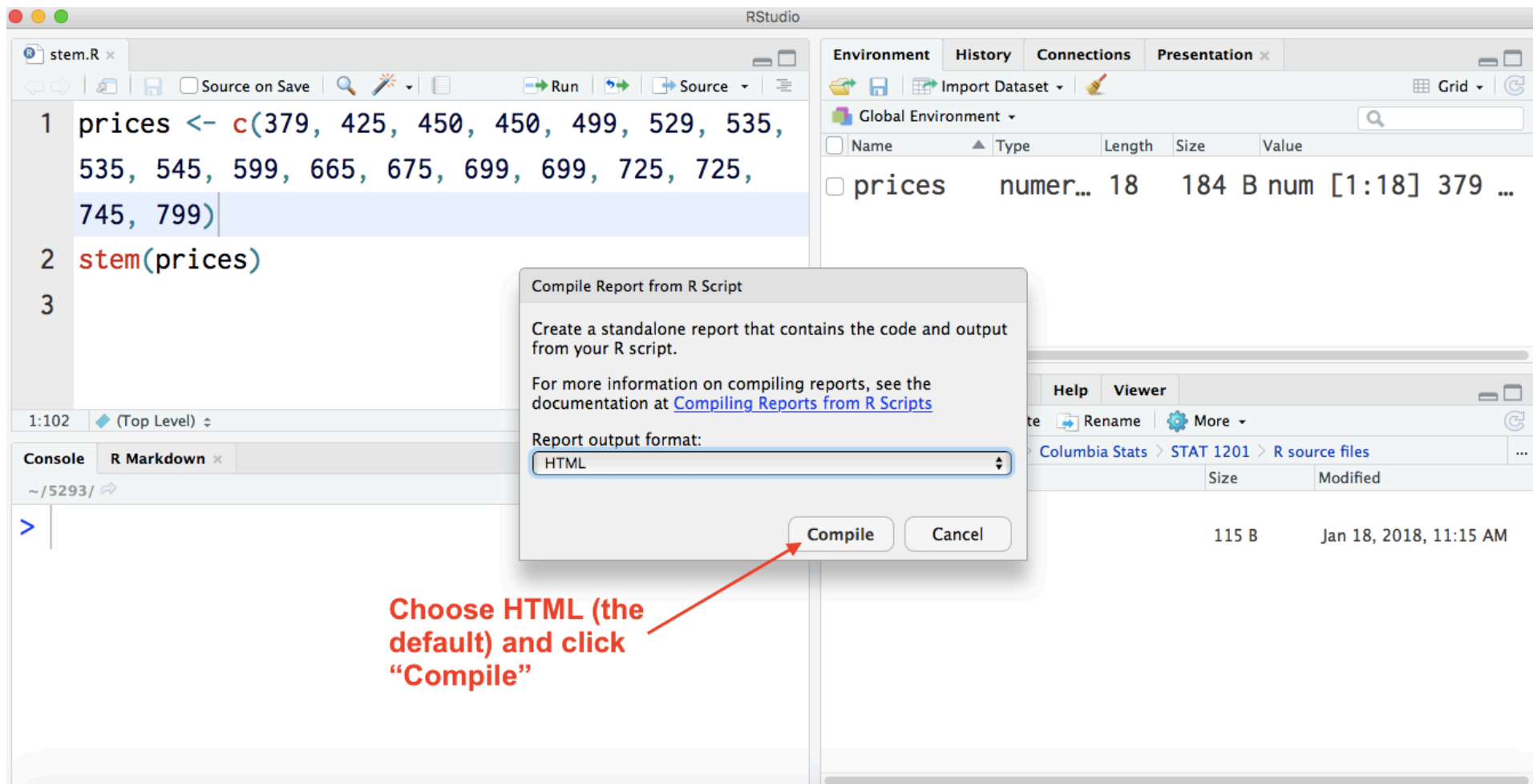
```
1 price <- c(425, 450, 450, 499, 529, 535, 535, 545, 599, 665, 675, 699, 699, 725, 725, 745, 799)
2 stem(prices)
```

A red box highlights the 'Save current document (⌘S)' button and the text: **If you work in the Source pane, you can save your work as .R files**  
**Run each line by moving the cursor to that line and pressing Command Enter**
- Environment Pane:** Shows the 'Global Environment' with a table of variables:

Name	Type	Length	Size	Value
prices	numer...	18	184 B	num [1:18] 379 ...
- Files Pane:** Shows the file explorer with the path: Home > Dropbox > Columbia Stats > STAT 1201 > R source files. It lists a file named '..' with a green arrow icon.
- Console:** Shows the prompt '>' and the path '~/5293/'.

# Compiling your work into an .html file





stem.R

```
1 prices <- c(379
535, 545, 599,
745, 799)
2 stem(prices)
3
```

1:102 (Top Level)

Console R Markdown

>

stem.html

Open in Browser

Find

Publish

stem.R

JoyceRobbins

Thu Jan 18 11:15:32 2018

prices <- c(379, 425, 450, 450, 499, 529, 535, 535, 545, 599, 665, 675, 699, 699, 725, 725, 745, 799)  
stem(prices)

##  
## The decimal point is 2 digit(s) to the right of the |  
##  
## 3 | 8  
## 4 | 355  
## 5 | 03445  
## 6 | 078  
## 7 | 00335  
## 8 | 0

Your compiled document,  
which shows code as well  
as output



RStudio

stem.R x

Source on Save Run Source

```
1 prices <- c(379, 425, 450, 450, 499, 529, 535,
2 535, 545, 599, 665, 675, 699, 699, 725, 725,
3 745, 799)
4 stem(prices)
```

1:102 (Top Level) R Script

Console R Markdown

~/5293/

Environment History Connections Presentation x

Import Dataset Grid

Global Environment

Name	Type	Length	Size	Value
prices	num...	18	184 B	num [1:18] 379 ...

Files Plots Packages Help Viewer

New Folder Delete Rename More


Home > Dropbox > Columbia Stats > STAT 1201 > R source files

Name	Size	Modified
..		
stem.R	115 B	Jan 18, 2018, 11:15 AM
stem.html	718.9 KB	Jan 18, 2018, 11:15 AM




An html file is automatically saved when you click Compile

1201/StemandLeafPlot.md at m · x

GitHub, Inc. [US] | https://github.com/jtr13/1201/blob/master/R/StemandLeafPlot.md

 This repository Search

Pull requestsIssuesMarketplaceExplore



jtr13 / 1201

Unwatch 4

Star 1

Fork 1

<> Code

Issues 0

Pull requests 0

Projects 0


Wiki

Insights

Settings

Branch: master 1201 / R / StemandLeafPlot.md

Find fileCopy path




 jtr13 Stem and leaf plots

214426b on Sep 17, 2017

1 contributor

19 lines (15 sloc) | 340 Bytes

RawBlameHistory




Stem and leaf plot

```
prices <- c(379, 425, 450, 450, 499, 529, 535, 535, 545, 599, 665,
           675, 699, 699, 725, 725, 745, 799)

stem(prices)

##
## The decimal point is 2 digit(s) to the right of the |
##
## 3 | 8
## 4 | 355
## 5 | 03445
## 6 | 078
## 7 | 00335
## 8 | 0
```

© 2018 GitHub, Inc. Terms Privacy Security Status Help



Contact GitHub API Training Shop Blog About

# Creating a stem-and-leaf plot

The screenshot shows the RStudio interface. In the Source editor, the first line of code is `stem(prices)`. The Console shows the output of this command, which is a stem-and-leaf plot. The plot has a title: "The decimal point is 2 digit(s) to the right of the |". The plot itself consists of a vertical line (the stem) with numbers to its left (the stems) and numbers to its right (the leaves). The stems are 3, 4, 5, 6, 7, and 8. The leaves are 8, 355, 03445, 078, 00335, and 0. A red text box explains that `stem()` is a function and that by "passing" the `prices` variable to the function, a stem-and-leaf plot is created.

```
> stem(prices)
```

The decimal point is 2 digit(s) to the right of the |

```
3 | 8
4 | 355
5 | 03445
6 | 078
7 | 00335
8 | 0
```

**stem() is a function. By "passing" the prices variable to the function, we get a stem and leaf plot of our data**

The Environment pane on the right shows the Global Environment with a table of variables. The table has columns for Name, Type, Length, Size, and Value. The variable `prices` is listed with Type `num`, Length `18`, Size `184 B`, and Value `num [1:18] 379 ...`.

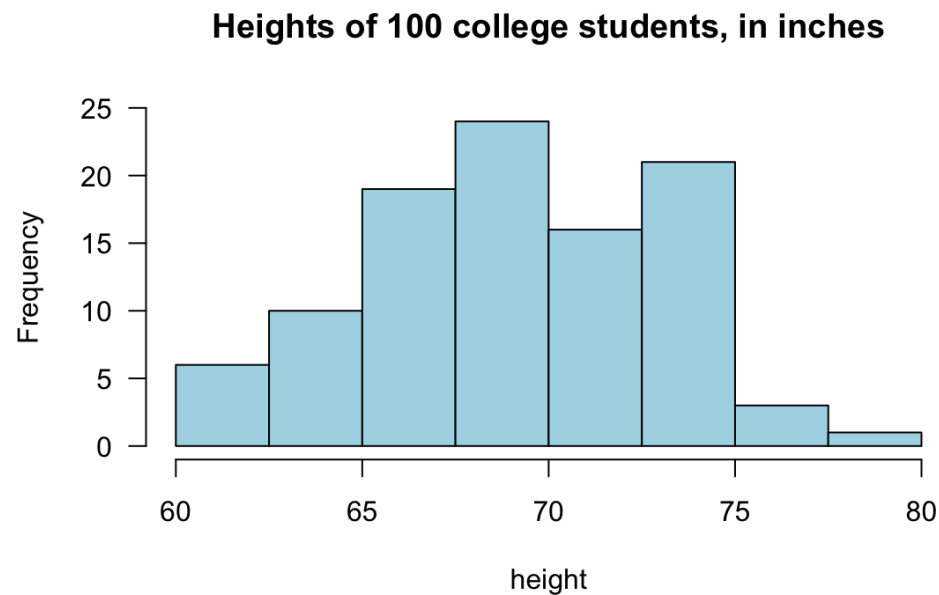
Name	Type	Length	Size	Value
prices	num	18	184 B	num [1:18] 379 ...

# Discrete data

```
## [1] "Heights of 100 college students, in inches"
```

```
##      [1] 60 60 61 61 61 62 63 63 64 64 64 64 65 65 65
##     [16] 65 66 66 66 66 66 66 67 67 67 67 67 67 67 67
##     [31] 67 67 67 67 67 68 68 68 68 68 68 68 69 69 69
##     [46] 69 69 69 69 69 69 70 70 70 70 70 70 70 70 71
##     [61] 71 71 71 71 71 72 72 72 72 72 72 72 72 72 72
##     [76] 73 73 73 74 74 74 74 74 74 74 74 74 74 74 74
##     [91] 74 75 75 75 75 75 75 76 76 77 79
```

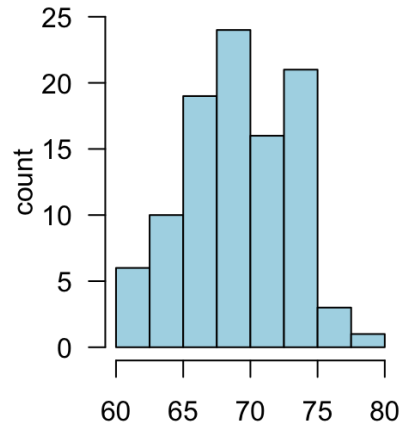
# Discrete data



```
##      [1] 60 60 61 61 61 62 63 63 64 64 64 64 65 65 65
##     [16] 65 66 66 66 66 66 66 66 67 67 67 67 67 67 67
##     [31] 67 67 67 67 67 68 68 68 68 68 68 68 68 69 69 69
##     [46] 69 69 69 69 69 69 70 70 70 70 70 70 70 70 71
##     [61] 71 71 71 71 71 72 72 72 72 72 72 72 72 72 72
##     [76] 73 73 73 74 74 74 74 74 74 74 74 74 74 74
##     [91] 74 75 75 75 75 75 76 76 77 79
```

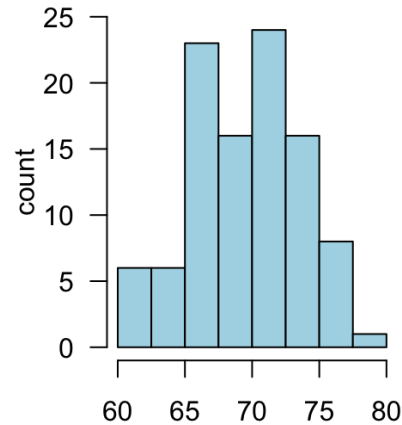
# Discrete data histogram

### Histogram of height



## RIGHT CLOSED, LEFT OPEN

### Histogram of height



**RIGHT OPEN, LEFT CLOSED**

##	[ 1]	60	60	61	61	61	62	63	63	64	64	64	64	65	65	65
##	[16]	65	66	66	66	66	66	66	67	67	67	67	67	67	67	67
##	[31]	67	67	67	67	67	68	68	68	68	68	68	68	69	69	69
##	[46]	69	69	69	69	69	69	70	70	70	70	70	70	70	70	71
##	[61]	71	71	71	71	71	72	72	72	72	72	72	72	72	72	72
##	[76]	73	73	73	74	74	74	74	74	74	74	74	74	74	74	74
##	[91]	74	75	75	75	75	75	76	76	77	79					

# EXERCISE

Draw a histogram of the asking prices for one-bedroom apartments in Morningside Heights (prices in thousands of \$)

Data source: cityrealty.com, 9/13/2016

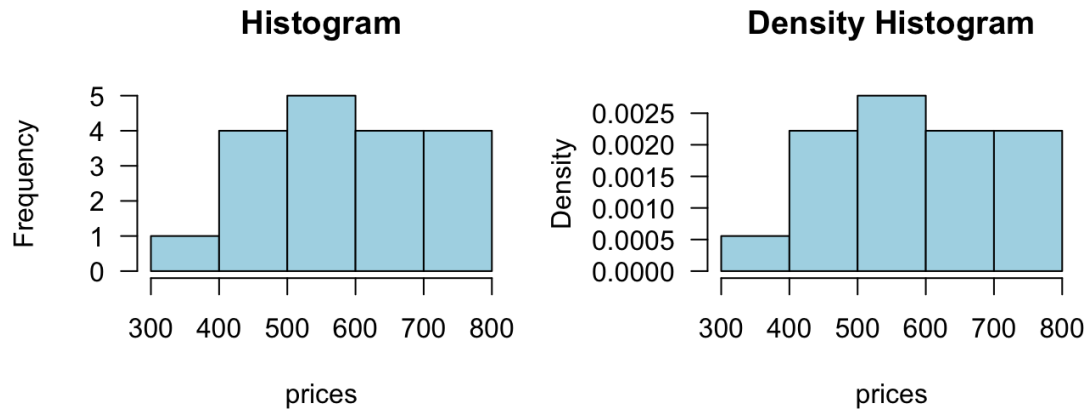
379, 425, 450, 450, 499, 529, 535, 535, 545,  
599, 665, 675, 699, 699, 725, 725, 745, 799

**Histogram of Morningside Heights  
One-Bedroom Apt. Prices**



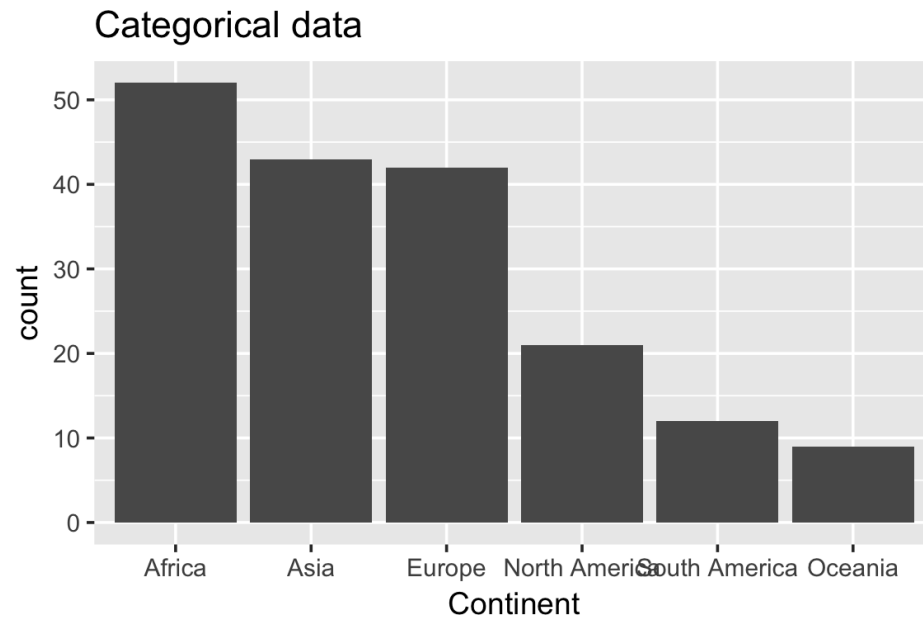


# Density histogram

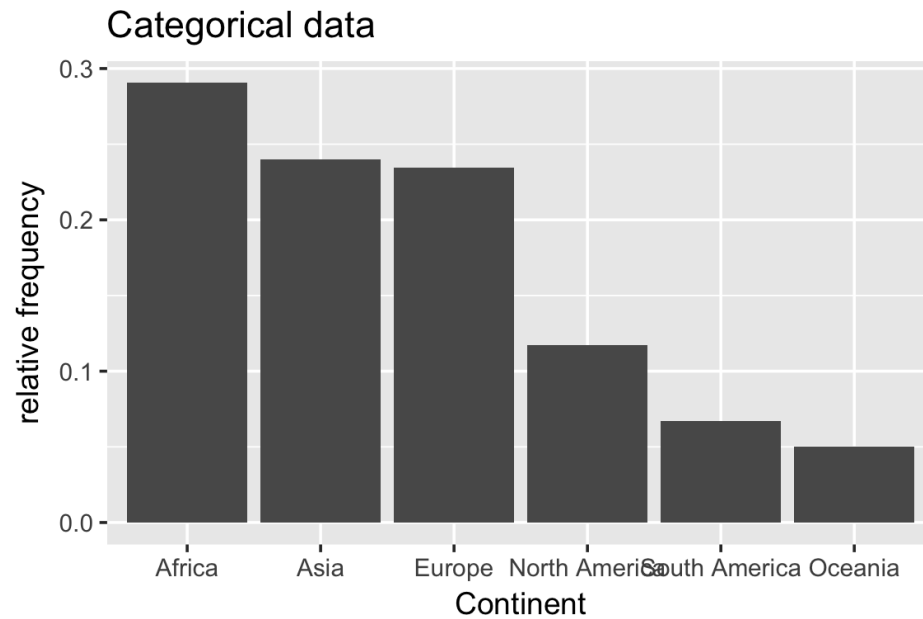


Class	Frequency	Rel. Frequency	Density
(300, 400]	1	.056	.00056
(400, 500]	4	.222	.00222
(500, 600]	5	.278	.00278
(600, 700]	4	.222	.00222
(700, 800]	4	.222	.00222

# Frequency bar chart



# Relative frequency bar chart



# Five number summary

1. min
2. lower fourth
3. median
4. upper fourth
5. max

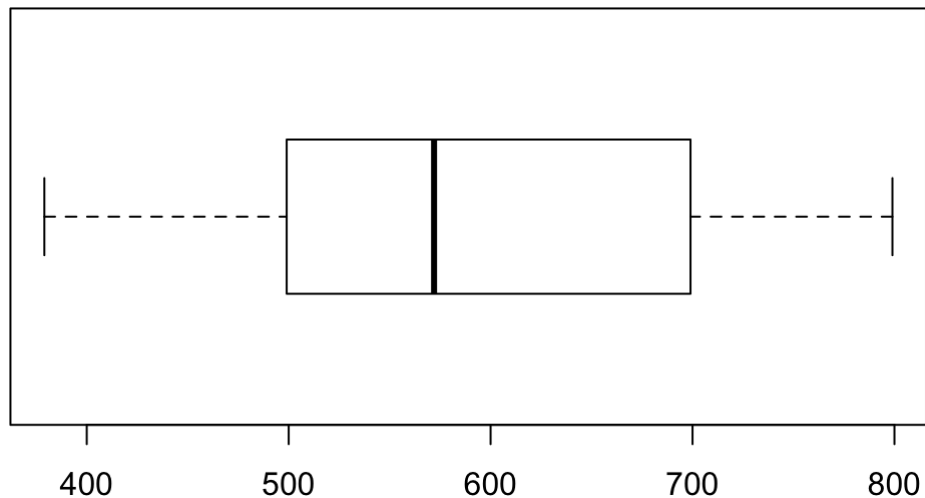
```
fivenum(prices)
```

```
## [1] 379 499 572 699 799
```

# Boxplot

379, 425, 450, 450, 499, 529, 535, 535, 545,  
599, 665, 675, 699, 699, 725, 725, 745, 799

```
## [1] 379 499 572 699 799
```



# Boxplot with outliers

