

# Big Data: Data Wrangling Boot Camp

## What is R?

Chuck Cartledge, PhD

23 February 2018

# Table of contents (1 of 1)

- 1 Intro.
- 2 What is R?
  - The language
- 3 RStudio
  - Basic how-tos (left side)
  - Basic how-tos (right side)
- 4 R Basics
  - Variables
- 5 Hands-on
- 6 Q & A
- 7 Conclusion
- 8 References

# What are we going to cover?

We're going to talk about:

- What is the language R?
- What GUI do I use to write and execute R programs?
- What are some basic variable types in R?



## The official definition.

*“R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.*

*R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, . . . ) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.”*

*CRAN Staff [2]*

# R is available for almost all major operating systems.

- Linux (and its variants)
- (Mac) OS X
- Windows



Get the R environment and a command line interface.

Download from: <https://cloud.r-project.org/>

Source code is available for custom OSs.

<https://github.com/wch/r-source>

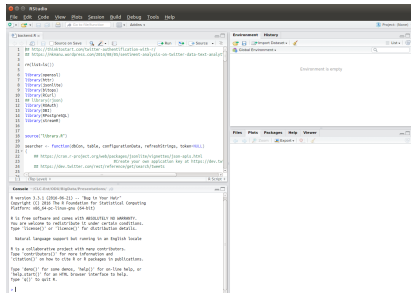
## Basic how-tos (left side)

# A complete IDE

A complete, integrated R development environment.

- ① Text editor
- ② R console
- ③ Variable list and contents
- ④ Tabbed display for different uses

See software overview and design document for version and download information.





Basic how-tos (left side)

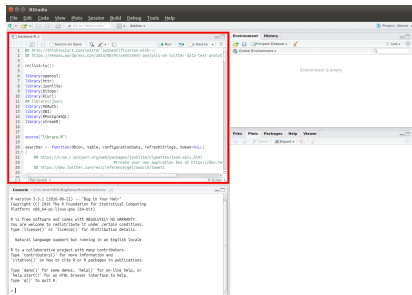
# Same image.

```

RStudio
File Edit View Plots Session Build Debug Tools Help
backlen.R
1 ## http://thelaststart.com/twitter-authentication-with-r/
2 ## https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bttops)
10 library(Rcurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OAuth)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Con, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   ## Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
27
28 # Top Level:
29
30 R Script
31
32 Console -/CLC-Env/ODU/BigData/Presentations/ /
33
34 R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
35 Copyright (C) 2016 The R Foundation for Statistical Computing
36 Platform: x86_64-pc-linux-gnu (64-bit)
37
38 R is free software and comes with ABSOLUTELY NO WARRANTY.
39 You are welcome to redistribute it under certain conditions.
40 Type 'license()' or 'licence()' for distribution details.
41
42 Natural language support but running in an English locale
43
44 R is a collaborative project with many contributors.
45 Type 'contributors()' for more information and
46 'citation()' on how to cite R or # packages in publications.
47
48 Type 'demo()' for some demos, 'help()' for on-line help, or
49 'help.start()' for an HTML browser interface to help.
50 Type 'q()' to quit R.
51
52 > |
  
```

# Editor

- “Smart” editor
- CTRL + O to open a file
- CTRL + S to save a file
- CTRL + A to highlight contents
- CTRL + Enter to transfer contents to Console
- Multiple files can be opened at once







Basic how-tos (left side)

# Same image.

```

1 ## http://thekostart.com/twitter-authentication-with-r/
2 ## https://mkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bitops)
10 library(RCurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OBS)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Con, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   ## Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
27
28 # Top Level: 3
  
```

```

Console ~CLC-Ent/ODU/BigData/Presentations /C/
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (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 # 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.

> |
  
```

# Console

- Interprets R commands
- Commands from editor, other panels, or manually entered
- Execution errors appear here
- Contents of print function appear here

The screenshot shows the RStudio interface with a script editor on the left and a console window at the bottom. The script editor contains R code for installing R and its dependencies. The console window shows the output of the R installation process, including the R version (3.5.1) and the R Foundation logo. The console output is highlighted with a red box.

```

1 #> http://www.r-project.org/
2 ## https://cran.r-project.org/web/packages/askpass/index.html#description
3
4 ##(ask-fs)
5
6 library(remotes)
7 library(rsp)
8 library(remotes)
9 library(rsp)
10 library(knitr)
11 ## library(rsp)
12 library(remotes)
13 library(askpass)
14 library(remotes)
15 library(remotes)
16
17 ##library("R")
18
19 searcher -> function@bin, base, configuration, refreshScripts, token@URL
20
21
22 ## https://cran.r-project.org/web/packages/askpass/index.html#description
23
24 ## https://www.r-project.org/web/packages/askpass/index.html#description
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```



Basic how-tos (left side)

# Same image.

The screenshot displays the RStudio environment. The script editor on the left contains the following code:

```

1 ## https://thekostart.com/twitter-authentication-with-r/
2 ## https://nkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bttops)
10 library(Rcurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OBS)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Conn, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   ## Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
27
28 # Top Level 2
29
30 R Script 2
  
```

The Environment pane on the right shows "Global Environment" and "Environment is empty".

The Console pane at the bottom, highlighted with a red border, shows the R startup message:

```

<CLC-End/ODDU/BigData/Presentations />
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (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 # packages in publications.

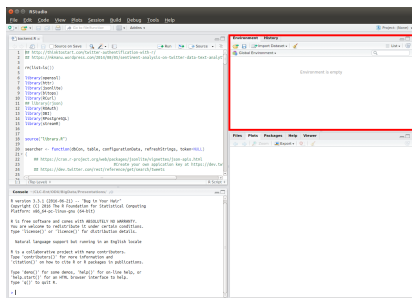
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
  
```



# Variables

- Displays contents of selected environment (including variables)
- Display history of console commands
- Can save and load data from data files





Basic how-tos (right side)

# Same image.

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for loading packages and defining a function. The code includes:
 

```
1 ## http://thelaststart.com/twitter-authentication-with-r/
2 ## https://nkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(openssl)
7 library(httr)
8 library(jsonlite)
9 library(bttops)
10 library(RCurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(OAuth)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher <- function(@Conn, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   ## Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
```
- Environment History:** Highlighted with a red box, showing "Environment is empty".
- Console:** Displays the R startup message:
 

```
R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (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

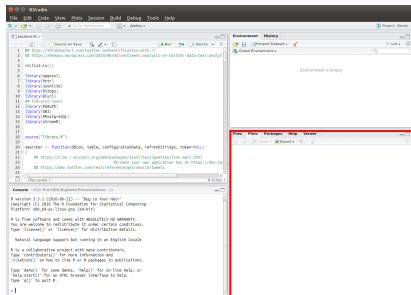
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or # packages in publications.

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.

-> |
```

# Tabbed display

- Displays files in the current directory
- Displays plots from the console
- Allows packages to be added, or removed from the console
- Provides help/man pages for R functions and packages





Basic how-tos (right side)

# Same image.

```

1 #> http://thelaststart.com/twitter-authentication-with-r/
2 #> https://nkmanu.wordpress.com/2014/08/05/sentiment-analysis-on-twitter-data-text-analyt
3
4 rm(list=la())
5
6 library(opensl)
7 library(httr)
8 library(jsonlite)
9 library(bitops)
10 library(RCurl)
11 ## library(rjson)
12 library(ROAuth)
13 library(DBI)
14 library(RPostgreSQL)
15 library(stream)
16
17
18 source("library.R")
19
20 searcher = function(@Conn, table, configurationData, refreshStrings, token=NULL)
21 {
22   ## https://cran.r-project.org/web/packages/jsonlite/vignettes/json-opts.html
23   #Create your own application key at https://dev.tw
24   ## https://dev.twitter.com/rest/reference/get/search/tweets
25
26 }
  
```

```

R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (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.

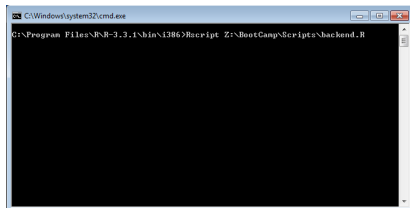
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
  
```

# Starting an R script in the background

The image shows a Windows environment.

A \*nix environment command is:  
`Rscript backend.R &`

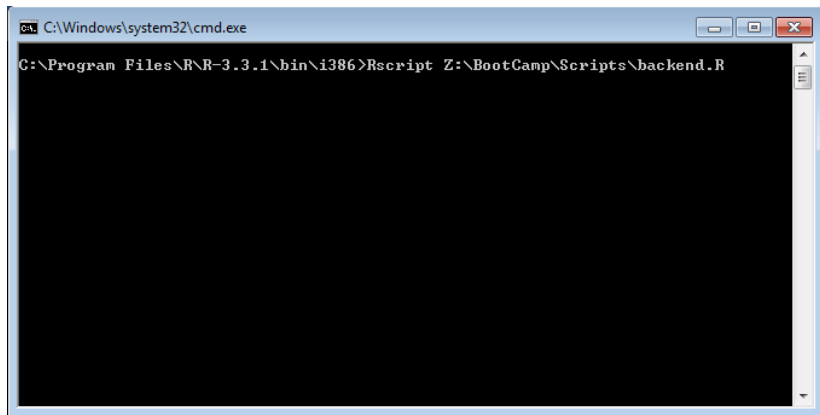


```
C:\Windows\system32\cmd.exe
G:\Program Files\R\R-3.3.1\bin\x86>Rscript Z:\BootCamp\Scripts\backend.R
```



Basic how-tos (right side)

# Same image.



```
C:\Windows\system32\cmd.exe  
C:\Program Files\R\R-3.3.1\bin\i386>Rscript Z:\BootCamp\Scripts\backend.R
```

# Basic help with functions[1]

## 1 Based on subject

```
1 help.search("data input")
```

## 2 Based on pattern matching

```
1 apropos("lm")
```

## 3 Looking for a specific item

```
1 find("lm")
```

## 4 About a specific item

```
1 ?lm
```

## 5 Example of a function

```
1 example(lm)
```

## 6 Source code for a function

```
1 lm
```

## 7 Demonstration of a function

```
1 demo(persp)
```

## 8 Demonstration of a function

```
1 vignette("moveline", package="
  grid")
```

## 9 Contents of a library

```
1 library(help=spatial)
```

## 10 Install a new library

```
1 install.packages("Kfn")
```

# Variable types (part 1 of 2)[3]

## 1 Variable names:

- Names are case sensitive
- Names cannot begin with numbers or special symbols
- Names cannot have internal spaces

## 2 Scalars (simple values)

```
1 variable <- 3
```

## 3 Vector of values

```
1 variable <- c(3, 5, 7, 10)
```

## 4 Matrix of values

```
1 variable <- matrix(1:20, nrow=5,
                    ncol=4)
```

## 5 Access matrix values

```
1 b <- variable[3,4]
2 b <- variable[,4]
3 b <- variable[3, ]
```

## 6 Array of values

```
1 variable <- array(1:20, dim=c
                  (2,2,5))
```

## Variable types (part 2 of 2)[3]

### 1 Data frames (each column must have the same number of values)

```
1 L3 <- LETTERS[1:3]
2 fac <- sample(L3, 10, replace = TRUE)
3 d <- data.frame(x = 1, y = 1:10, fac = fac)
```

### 2 Lists (each list entry can have a different number of values)

```
1 pts <- list(x = cars[,1], y = cars[1:23,2])
```

### 3 Saving variables

```
1 save(x, y, file = "xy.RData")
```

### 4 Loading variables

```
1 load(file = "xy.RData")
```

# Functions are supported

- 1 Have the same naming conventions as variables
- 2 Have three parts:
  - 1 Optional pass parameters (named, evaluated, unnamed)
  - 2 Text of the function
  - 3 The environment where and while the function executes

```
1 functionName <- function(p1, p2=1234, ...)  
2 {  
3   p3 <- p1 + p2  
4   return(p3)  
5 }  
6  
7 functionName(3)  
8 1237  
9  
10 functionName(3, 6677)  
11 6680
```

- 3 The last value evaluated is returned.
- 4 Statements grouped by "curly braces" or semicolons.

# Some simple exercises to get familiar with R and RStudio

- 1 Create a variable and assign it the value 3
- 2 Print your variable
- 3 Create a function that takes one parameter and returns the square of that value
- 4 Use your function to compute the square of 45
- 5 Print the value of the passed parameter inside the function
- 6 Open the file `library.R` and explain what the function `dumpObject` does

## Q & A time.

Q: Why was Stonehenge  
abandoned?

A: It wasn't IBM compatible.



## What have we covered?

- Covered a little bit of R's background
- Looked at RStudio, a cross platform GUI for working with R
- Looked at some R basics (variable types and functions)



Next: Look at some of R's built-in datasets and specifically; the iris dataset



## References (1 of 1)

- [1] Michael J. Crawley, [The R Book](#), John Wiley & Sons, 2012.
- [2] CRAN Staff, [What is R?](#),  
<https://www.r-project.org/about.html>, 2017.
- [3] Simon Walkowiak, [Big Data Analytics with R](#), Packt Publishing Ltd., 2016.