****Chapter 1****

1. <https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/datasets/women.csv>.
2. <https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/slr/excel/slr01.xls>
3. <https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/datasets/chickwts.csv>

****Chapter 2****

1. <https://archive.ics.uci.edu/ml/machine-learning-databases/wine/>
2. <https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/datasets/USArrests.csv>
3. ML databases – bridges in UCI archive
4. [https://www.kaggle.com/datasets/heitornunes/pittsburgh-bridges-data-set](https://protect-eu.mimecast.com/s/jxv2CAPqquOEJKpCGod0-?domain=kaggle.com)

Unicef data: Under-five mortality rate since 1990

<https://data.unicef.org/topic/child-survival/under-five-mortality/>

****Chapter 3****

1. <https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/slr/excel/slr12.xls>
2. <https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/slr/excel/slr05.xls>
3. <https://people.sc.fsu.edu/~jburkardt/datasets/martinez/martinez.html>
4. <https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/slr/excel/slr07.xls>
5. <https://archive.ics.uci.edu/ml/machine-learning-databases/00436/Container_Crane_Controller_Data_Set.csv>
6. <https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/DAAG/allbacks.csv>
7. GMC truck datasets

<http://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/slr/excel/slr01.xls>

****Chapter 4****

1. housing.txt (53MB)
2. nyc-deaths.csv (4KB)
3. <https://data.unicef.org/topic/water-and-sanitation/hygiene/drinking-water-sanitation-hygiene-database-July-2017.xlsx> (456 KB)
4. <http://apps.who.int/gho/data/node.main.488?lang=en> Medicine.csv (1.2 KB)

****Chapter 5****

1. <http://vincentarelbundock.github.io/Rdatasets/csv/datasets/longley.csv>

1. <http://www-bcf.usc.edu/~gareth/ISL/Advertising.csv>
2. All Greens Franchise Data Set

<https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/mlr/excel/mlr05.xls>

1. <https://gist.githubusercontent.com/tijptjik/9408623/raw/b237fa5848349a14a14e5d4107dc7897c21951f5/wine.csv>

UCI Wheat Data Set

<http://archive.ics.uci.edu/ml/machine-learning-databases/00236/seeds_dataset.txt>

UCI Travel Reviews Data Set

<https://archive.ics.uci.edu/ml/machine-learning-databases/00484/tripadvisor_review.csv>

1. boston.csv [35KB]
2. <https://archive.ics.uci.edu/ml/machine-learning-databases/00451/dataR2.csv> Description: UCI Machine Learning Repository: Breast Cancer Coimbra Dataset; cancer.csv [7.5KB]

****Chapter 6****

GitHub with code from “Practical Data Science with R” book

<https://github.com/WinVector/zmPDSwR/tree/master/Custdata>

stocks.csv [1KB]

Cloth dataset:

Bissell, A.F. (1972) A negative binomial model with varying element size. Biometrika, 59,

435–441

<https://vincentarelbundock.github.io/Rdatasets/datasets.html>

size.csv [1KB]

Beaver dataset taken from:

P. S. Reynolds (1994) Time-series analyses of beaver body temperatures. Chapter 11 of Lange,

N., Ryan, L., Billard, L., Brillinger, D., Conquest, L. and Greenhouse, J. eds (1994) *Case Studies in*

*Biometry.* New York: John Wiley and Sons

The dataset is available from:

<https://vincentarelbundock.github.io/Rdatasets/csv/boot/beaver.csv>

1. dating.csv [36KB]
2. iqsize.csv [1KB]

****Chapter 7****

1. MySQL tutorial page at http://www.mysqltutorial.org/mysql-sample-database.aspx. Download the zipped mysqlsampledatabase.zip
2. world.sql [397 KB]
3. auto.sql [22 KB]

Heterogeneous Integrated Dataset for Maritime Intelligence, Surveillance, and Reconnaissance available at https://zenodo.org/record/1167595/files/%5BP1%5D%20AIS%20Data.zip?download=1

nari\_dynamic\_sar.csv [221 KB]

****Chapter 8****

1. regression.csv [1KB]
2. Australian Journal of Zoology, Vol. 28, p607-613). The dataset can be downloaded from (<https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/slr/excel/slr07.xls>
3. ratings.csv [1KB]

Airline Costs as Function of 7 Operating Variables from the Statistics department of University of Florida: http://users.stat.ufl.edu/~winner/data/airline\_costs.dat

1. rock.csv [2KB]
2. Conventional and Social Media Movies dataset from UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/machine-learning-databases/00424/2014%20and%202015%20CSM%20dataset.xlsx>

****Chapter 9****

Titanic: Machine Learning from Disaster

<https://www.kaggle.com/c/titanic>

Social media ads data

<https://www.kaggle.com/rakeshrau/social-network-ads>

Github dataset for hsbdemo

<https://github.com/rlowrance/re/blob/master/hsbdemo.csv>

Car evaluation dataset

<http://archive.ics.uci.edu/ml/machine-learning-databases/car/car.data>

Wikipedia iris flower data set

<https://en.wikipedia.org/wiki/Iris_flower_data_set>

Balloons dataset

<https://archive.ics.uci.edu/ml/machine-learning-databases/balloons/adult+stretch.data>

Balloons dataset yellow small

<https://archive.ics.uci.edu/ml/machine-learning-databases/balloons/yellow-small.data>

Lenses dataset

<http://archive.ics.uci.edu/ml/machine-learning-databases/lenses/lenses.data>

UCI Bank marketing data set

<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

1. golf.csv [1KB]
2. regression.csv [1KB]

Combined cycle power plant dataset

<http://archive.ics.uci.edu/ml/machine-learning-databases/00294/CCPP.zip>

crash.csv [1KB]

quality.csv [1KB]

Immunotherapy dataset used by F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, A. Khosravi, P. Layegh, and S. Nahavandi, “An expert system for selecting wart treatment method,” Computers in Biology and Medicine, vol. 81, pp. 167-175, 2/1/ 2017. Available from UCI machine learning repository at: <https://archive.ics.uci.edu/ml/machine-learning-databases/00428/Immunotherapy.xlsx>

iris.csv [5KB]

Horseshoe crab dataset is available in R studio version 3.4.4 and also can be downloaded from outside University of Florida statistics department’s repository: <http://users.stat.ufl.edu/~aa/cda/data/Crabs.dat>

weather.csv [1KB]

Office of Population Research, Princeton University, 1976. Unpublished data assembled under NICHD contract number No 1-HD-O-2077.” It is available at this link (<https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/datasets/swiss.csv>).

NFL 2014 Combine Performance Results data sourced from nfl.com and available from University of Florida at: <http://users.stat.ufl.edu/~winner/data/nfl_combine_2014.dat>

overdrawn.csv [7KB]

tipjoke.csv [4KB]

UCI machine learning repository

<http://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data>

Blues Guitarists Hand Posture and Thumbing Style by Region and Birth Period from A.M. Cohen (1996). "The Hands of Blues Guitarists," American Music, Vol. 14, #4, pp. 455-479 and available at: <http://users.stat.ufl.edu/~winner/data/blues_hand.dat>

1. <http://archive.ics.uci.edu/ml/machine-learning-databases/00380/YouTube-Spam-Collection-v1.zip>
2. German Teachers and Nazi party membership data sourced from K.H. Jarausch and G. Arminger (1989),   
   “The German Teaching Profession and Nazi Party Membership: A Demographic Logit Model”, Journal of Interdisciplinary History, Vol. 20, pp 197-225. The dataset is available from University of Florida at: <http://users.stat.ufl.edu/~winner/data/nazi.dat>
3. <http://stat.pugetsound.edu/hoard/datasets/anthrokids.csv>

Portuguese sea battles data sourced from Armando da Silva Saturnino Monteiro (2001). “Decline and Fall of Portuguese Seapower 1583-1663”, Journal of Military History Vol. 65, #1, pp9-20. The dataset is available at: <http://users.stat.ufl.edu/~winner/data/armada.dat>

****Chapter 10****

1. StoneFlakes data set

<https://archive.ics.uci.edu/ml/machine-learning-databases/00299/StoneFlakes.dat>

1. User knowledge modeling dataset

<https://archive.ics.uci.edu/ml/machine-learning-databases/00257/Data_User_Modeling_Dataset_Hamdi%20Tolga%20KAHRAMAN.xls>

1. lifecyclesaving.csv [2KB]
2. 2008 season statistics for Women's professional golfers on LPGA tour, sourced from lgpa.com and available from University of Florida: <http://users.stat.ufl.edu/~winner/data/lpga2008.dat>
3. University of California, Irvine Machine Learning Repository

<http://archive.ics.uci.edu/ml/machine-learning-databases/concrete/slump/slump_test.data>

SGEMM GPU kernel performance data set created by Enrique G. Paredes (egparedes@ifi.uzh.ch) and Rafael Ballester-Ripoll ([rballester@ifi.uzh.ch](mailto:rballester@ifi.uzh.ch)) and available from UCI machine learning repository at: <http://archive.ics.uci.edu/ml/machine-learning-databases/00440/sgemm_product_dataset.zip>

****Chapter 11****

1. dermatology.csv [26 KB]
2. twitter\_search.zip [78 KB]
3. youtube\_search.py [4 KB]
4. json2csv\_user.py [1 KB]
5. json2csv\_business.py [1 KB]
6. yelp\_academic\_dataset\_user.json.csv [120 MB]
7. yelp\_academic\_dataset\_business.json.csv [27 MB]
8. hatecrime.csv [4 KB]