

Week 1

Contents

1	Week 1	2
1.1	Obtaining Data Motivation	2
1.2	Raw and Processed Data	2
1.3	Components of Tidy Data	2
1.3.1	The four things you should have	2
1.3.2	The raw data	2
1.3.3	The tidy data	3
1.3.4	The code book	3
1.3.5	The instruction list	3
1.4	Downloading files	3
1.4.1	Get/set your working directory	3
1.4.2	Checking for and creating directories	3
1.4.3	Getting data from the internet - <code>download.file()</code>	4
1.4.4	Example - Baltimore camera data	4
1.4.5	Some notes about <code>download.file()</code>	4
1.5	Reading local flat files	4
1.5.1	Loading flat files - <code>read.tableO</code>	4
1.5.2	Some more important parameters	5
1.6	Reading Excel Files	5
1.6.1	<code>read.xlsx()</code> , <code>read.xlsx2()</code> , {xlsx package}	5
1.6.2	Reading specific rows and columns	5
1.6.3	Further notes	5
1.7	Reading XML	6
1.7.1	XML	6
1.7.2	Tags, elements and attributes	6
1.7.3	Example XML file	6
1.7.4	Directly access parts of the XML document	8
1.7.5	Programaticallye xtractp artso f the file	8
1.7.6	XPath	8
1.7.7	Get the items on the menu and prices	8
1.7.8	Another example	9
1.7.9	Viewing the source	9
1.7.10	Extract content by attributes	9
1.8	Reading JSON	10
1.8.1	JSON	10
1.8.2	Example JSON file	10
1.8.3	Writing data frames to JSON	11
1.8.4	Convert back to JSON	30
1.8.5	Further resources	31
1.9	The <code>data.table</code> Package	31
1.9.1	<code>data.table</code>	31
1.9.2	Create data tables just like data frames	31
1.9.3	See all the data tables in memory	31
1.9.4	Subsetting rows	32

1.9.5	Subsetting columns!?	32
1.9.6	Column subsetting in <code>data.table</code>	32
1.9.7	Calculating values or variables with expressions	33
1.9.8	Adding new columns	33
1.9.9	Multiple operations	34
1.9.10	plyr like operations	34
1.9.11	Special variables	34
1.9.12	Keys	35
1.9.13	Joins	37
1.9.14	Fast reading	38
1.9.15	Summary and further reading	38

1 Week 1

1.1 Obtaining Data Motivation

Contents:

- Finding and extracting raw data
- Tidy data principles and how to make tidy data
- Practical implementation through a range of R packages

Purpose: Raw data -> Processing script -> tidy data -> data analysis -> data communication.

1.2 Raw and Processed Data

Qualitative: Country of origin, sex, treatment Quantitative: Height, weight, blood pressure

- Raw data:
 - The original source of data
 - hard to use
 - data analysis includes processing
 - raw data may only need to be processed once
- Processed data:
 - ready for analysis
 - processing can include merging, subsetting, transforming, etc
 - there may be standards for processing all steps should be recorded

1.3 Components of Tidy Data

This is the target.

1.3.1 The four things you should have

1. The *raw data*.
2. A *tidy data set*
3. A *code book* describing each variable and its values in the tidy data set.
4. An explicit and exact *recipe* you used to go from 1 -> 2,3.

1.3.2 The raw data

- The strange binary file your measurement machine spits out
- The unformatted Excel file with 10 worksheets the company you contracted with sent you
- The complicated JSON data you got from scraping the Twitter API
- The hand-entered numbers you collected looking through a microscope

You know the raw data is in the right format if you 1. Ran no software on the data 2. Did not manipulate any of the numbers in the data 3. You did not remove any data from the data set 4. You did not summarize the data in any way

<https://github.com/jtleek/datasharing>

1.3.3 The tidy data

1. *Each variable* you measure should be in one *column*
2. *Each different observation* of that variable should be in a different *row*
3. There should be one *table for each “kind” of variable*
4. If you have multiple tables, they should include a column in the table that *allows them to be linked*

Some other important tips Include a row at the top of each file with variable names. Make variable names *human readable* AgeAtDiagnosis instead of AgeDx In general data should be saved in *one file per table*.

1.3.4 The code book

1. Information about the variables (including units!) in the data set not contained in the tidy data
 2. Information about the summary choices you made
 3. Information about the experimental study design you used
- Some other important tips A common format for this document is a Word/text file. There should be a section called “Study design” that has a thorough description of how you collected the data. There must be a section called “Code book” that describes each variable and its units.

1.3.5 The instruction list

Ideally a computer script (in R :-), but I suppose Python is ok too ...) The input for the script is the raw data The output is the processed, tidy data There are no parameters to the script

In some cases it will not be possible to script every step. In that case you should provide instructions like:

1. Step 1 - take the raw file, run version 3.1.2 of summarize software with parameters a=1, b=2, C=3
2. Step 2 - run the software separately for each sample
3. Step 3 - take column three of outputfile.out for each sample and that is the corresponding row in the output data set

1.4 Downloading files

1.4.1 Get/set your working directory

A basic component of working with data is knowing your working directory The two main commands are `getwd ()` and `setwd ()`. Be aware of relative versus absolute paths - Relative - Absolute

Important difference in Windows `setwd (c : \\users\\ \\Andrew\\ \\Downloads 11)`

1.4.2 Checking for and creating directories

`file.exists` `dir.create` Here is an example checking for a “data” directory and creating it if it doesn’t exist

```
file.exists("data")
```

```
## [1] TRUE
```

```
if(!file.exists("data")){  
  dir.create("data")  
}
```

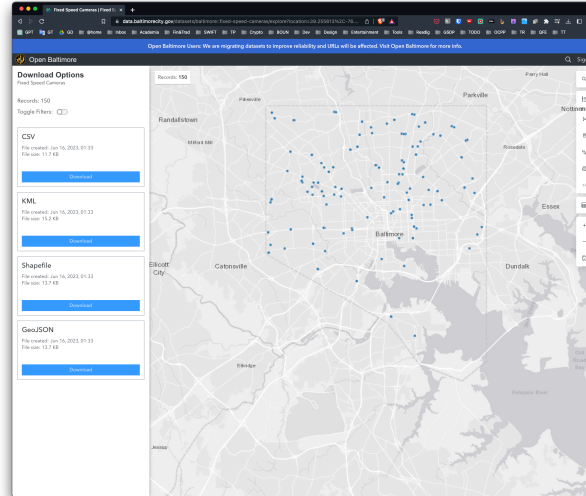
1.4.3 Getting data from the internet - download.file()

Downloads a file from the internet Even if you could do this by hand, helps with reproducibility Important parameters are url, destfile, method Useful for downloading tab-delimited, csv, and other files

```
str(download.file())
```

```
## function (url, destfile, method, quiet = FALSE, mode = "w", cacheOK = TRUE,  
##      extra = getOption("download.file.extra"), headers = NULL, ...)
```

1.4.4 Example - Baltimore camera data



```
fileUrl <- "https://opendata.arcgis.com/api/v3/datasets/cc4d3f4c436d4736b121ce781d4f86de_0/downloads/data  
download.file(fileUrl, destfile = "./data/cameras.csv", method = "curl")  
list.files("./data")
```

```
## [1] "cameras.csv" "simple.xml"
```

1.4.5 Some notes about download.file()

If the url starts with http you can use download.file() If the url starts with https on Windows you may be ok If the url starts with https on Mac you may need to set method="curl" If the file is big, this might take a while Be sure to record when you downloaded.

1.5 Reading local flat files

1.5.1 Loading flat files - read.table()

This is the main function for reading data into R Flexible and robust but requires more parameters Reads the data into RAM - big data can cause problems Important parameters file, header, sep, row.names, nrows Related: read.csv(), read.csv2()

```
cameraData <- read.table("./data/cameras.csv")
```

```
## Error in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, : line 1 did not have 2
```

```
cameraData <- read.table("data/cameras.csv", sep = ",", header = TRUE)
```

```
## Error in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, : line 37 did not have 3
```

apparently one or more rows has different number of columns.

1.5.2 Some more important parameters

quote - you can tell R whether there are any quoted values quote="" means no quotes. na.strings - set the character that represents a missing value. nrows - how many rows to read of the file (e.g. nrows=10 reads 10 lines). skip- number of lines to skip before starting to read

In my experience, the biggest trouble with reading flat files are quotation marks ' or " placed in data values, setting quote="" often resolves these.

```
str(read.table)

## function (file, header = FALSE, sep = "", quote = "\"'", dec = ".", numerals = c("allow.loss",
##   "warn.loss", "no.loss"), row.names, col.names, as.is = !stringsAsFactors,
##   tryLogical = TRUE, na.strings = "NA", colClasses = NA, nrows = -1,
##   skip = 0, check.names = TRUE, fill = !blank.lines.skip, strip.white = FALSE,
##   blank.lines.skip = TRUE, comment.char = "#", allowEscapes = FALSE,
##   flush = FALSE, stringsAsFactors = FALSE, fileEncoding = "", encoding = "unknown",
##   text, skipNul = FALSE)
```

1.6 Reading Excel Files

1.6.1 read.xlsx(), read.xlsx2(), {xlsx package}

```
library(xlsx)

## Warning in system("/usr/libexec/java_home", intern = TRUE): running command
## '/usr/libexec/java_home' had status 1

## Warning in fun(libname, pkgname): Cannot find JVM library 'NA/lib/server/libjvm.dylib'
## Install Java and/or check JAVA_HOME (if in doubt, do NOT set it, it will be detected)

## Error: package or namespace load failed for 'xlsx':
## .onLoad failed in loadNamespace() for 'rJava', details:
##   call: fun(libname, pkgname)
##   error: JVM could not be found
```

1.6.2 Reading specific rows and columns

```
colIndex <- 2:3
rowIndex <- 1:4
cameraDataSubset <-
  read.xlsx(
    "data/data.xlsx",
    sheetIndex = 1,
    colIndex = colIndex,
    rowIndex = rowIndex
  )

## Error in read.xlsx("data/data.xlsx", sheetIndex = 1, colIndex = colIndex, : could not find function
```

1.6.3 Further notes

- The write.xlsx function will write out an Excel file with similar arguments.
- read.xlsx2 is much faster than read.xlsx but for reading subsets of rows may be slightly unstable.
- XLConnect package has more options for writing and manipulating Excel files.
- XLConnect vignette is a good place to start for that package
- In general it is advised to store your data in either a database or in a comma separated files (.csv) or tab separated files (.tab/.txt) as they are easier to distribute.

1.7 Reading XML

1.7.1 XML

Extensible markup language Frequently used to store structured data Particularly widely used in internet applications Extracting XML is the basis for most web scraping Components - Markup - labels that give the text structure - Content - the actual text of the document

1.7.2 Tags, elements and attributes

Tags correspond to general labels - Start tags - End tags - Empty tags

Elements are specific examples of tags

- Hello, World

Attributes are components of the label

-
- Connect A to B.

1.7.3 Example XML file

```
library(XML)
xmlURL <- "https://www.w3schools.com/xml/simple.xml"
doc <- xmlTreeParse(xmlURL, useInternal = TRUE)
```

```
## Error: XML content does not seem to be XML: ''
```

```
xmlURL <- "https://www.w3schools.com/xml/simple.xml"
download.file(xmlURL, destfile = "./data/simple.xml", method = "curl")
list.files("./data")
```

```
## [1] "cameras.csv" "simple.xml"
```

```
library(XML)
# Parse the XML file
doc <- xmlTreeParse("data/simple.xml", useInternal = TRUE)
doc
```

```
## <?xml version="1.0" encoding="UTF-8"?>
## <breakfast_menu>
##   <food>
##     <name>Belgian Waffles</name>
##     <price>$5.95</price>
##     <description>Two of our famous Belgian Waffles with plenty of real maple syrup</description>
##     <calories>650</calories>
##   </food>
##   <food>
##     <name>Strawberry Belgian Waffles</name>
##     <price>$7.95</price>
##     <description>Light Belgian waffles covered with strawberries and whipped cream</description>
##     <calories>900</calories>
##   </food>
##   <food>
##     <name>Berry-Berry Belgian Waffles</name>
##     <price>$8.95</price>
##     <description>Light Belgian waffles covered with an assortment of fresh berries and whipped cream</description>
##     <calories>900</calories>
```

```

## </food>
## <food>
##   <name>French Toast</name>
##   <price>$4.50</price>
##   <description>Thick slices made from our homemade sourdough bread</description>
##   <calories>600</calories>
## </food>
## <food>
##   <name>Homestyle Breakfast</name>
##   <price>$6.95</price>
##   <description>Two eggs, bacon or sausage, toast, and our ever-popular hash browns</description>
##   <calories>950</calories>
## </food>
## </breakfast_menu>
##

rootNode <- xmlRoot(doc)
rootNode

## <breakfast_menu>
##   <food>
##     <name>Belgian Waffles</name>
##     <price>$5.95</price>
##     <description>Two of our famous Belgian Waffles with plenty of real maple syrup</description>
##     <calories>650</calories>
##   </food>
##   <food>
##     <name>Strawberry Belgian Waffles</name>
##     <price>$7.95</price>
##     <description>Light Belgian waffles covered with strawberries and whipped cream</description>
##     <calories>900</calories>
##   </food>
##   <food>
##     <name>Berry-Berry Belgian Waffles</name>
##     <price>$8.95</price>
##     <description>Light Belgian waffles covered with an assortment of fresh berries and whipped cream</description>
##     <calories>900</calories>
##   </food>
##   <food>
##     <name>French Toast</name>
##     <price>$4.50</price>
##     <description>Thick slices made from our homemade sourdough bread</description>
##     <calories>600</calories>
##   </food>
##   <food>
##     <name>Homestyle Breakfast</name>
##     <price>$6.95</price>
##     <description>Two eggs, bacon or sausage, toast, and our ever-popular hash browns</description>
##     <calories>950</calories>
##   </food>
## </breakfast_menu>

xmlName(rootNode)

## [1] "breakfast_menu"

```

```
names(rootNode)
```

```
## food food food food food
## "food" "food" "food" "food" "food"
```

1.7.4 Directly access parts of the XML document

```
rootNode[[1]]
```

```
## <food>
## <name>Belgian Waffles</name>
## <price>$5.95</price>
## <description>Two of our famous Belgian Waffles with plenty of real maple syrup</description>
## <calories>650</calories>
## </food>
```

```
rootNode[[1]][[1]]
```

```
## <name>Belgian Waffles</name>
```

1.7.5 Programaticallye xtractp artso f the file

```
xmlSApply(rootNode,xmlValue)
```

```
##
## "Belgian Waffles$5.95Two of our famous Belgian Waffles with plenty of 1
##
## "Strawberry Belgian Waffles$7.95Light Belgian waffles covered with strawberries and
##
## "Berry-Berry Belgian Waffles$8.95Light Belgian waffles covered with an assortment of fresh berries and
##
## "French Toast$4.50Thick slices made from our homemade
##
## "Homestyle Breakfast$6.95Two eggs, bacon or sausage, toast, and our ever-pop
```

1.7.6 XPath

- /node Top level node
- //node Node at any level
- node[@attr-name] Node with an attribute name
- node[@attr-name='bob'] Node with attribute name attr-name='bob'

1.7.7 Get the items on the menu and prices

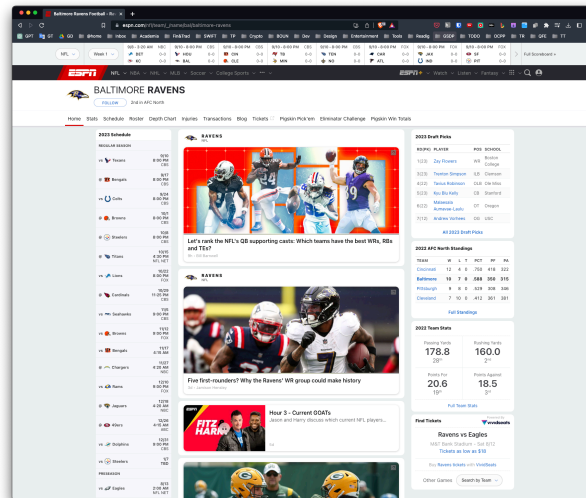
```
xpathSApply(rootNode, "//name",xmlValue)
```

```
## [1] "Belgian Waffles" "Strawberry Belgian Waffles"
## [3] "Berry-Berry Belgian Waffles" "French Toast"
## [5] "Homestyle Breakfast"
```

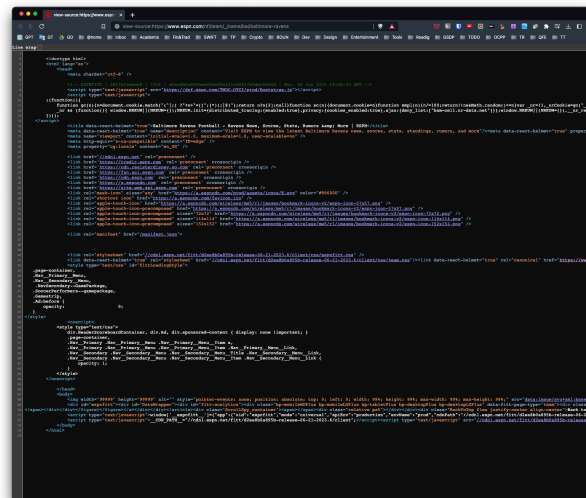
```
xpathSApply(rootNode, "//price",xmlValue)
```

```
## [1] "$5.95" "$7.95" "$8.95" "$4.50" "$6.95"
```


1.7.8 Another example



1.7.9 Viewing the source



1.7.10 Extract content by attributes

[illegible]

1.8 Reading JSON

1.8.1 JSON

Javascript Object Notation Lightweight data storage Common format for data from application programming interfaces (APIs) Similar structure to XML but different syntax/format Data stored as - Numbers (double) - Strings (double quoted) - Boolean (true or false) - Array (ordered, comma separated enclosed in square brackets[]) - Object (unorderd, comma separated collection of key:value pairs in curly brackets {})

<http://en.wikipedia.org/wiki/JSON>

1.8.2 Example JSON file

```
repoURL <- "https://api.github.com/users/jtleek/repos"
library(jsonlite)
jsonData <- fromJSON(repoURL)
names(jsonData)
```

```
## [1] "id" "node_id"
## [3] "name" "full_name"
## [5] "private" "owner"
## [7] "html_url" "description"
## [9] "fork" "url"
## [11] "forks_url" "keys_url"
## [13] "collaborators_url" "teams_url"
## [15] "hooks_url" "issue_events_url"
## [17] "events_url" "assignees_url"
## [19] "branches_url" "tags_url"
## [21] "blobs_url" "git_tags_url"
## [23] "git_refs_url" "trees_url"
## [25] "statuses_url" "languages_url"
## [27] "stargazers_url" "contributors_url"
## [29] "subscribers_url" "subscription_url"
## [31] "commits_url" "git_commits_url"
## [33] "comments_url" "issue_comment_url"
## [35] "contents_url" "compare_url"
## [37] "merges_url" "archive_url"
## [39] "downloads_url" "issues_url"
## [41] "pulls_url" "milestones_url"
## [43] "notifications_url" "labels_url"
## [45] "releases_url" "deployments_url"
## [47] "created_at" "updated_at"
## [49] "pushed_at" "git_url"
## [51] "ssh_url" "clone_url"
## [53] "svn_url" "homepage"
## [55] "size" "stargazers_count"
## [57] "watchers_count" "language"
## [59] "has_issues" "has_projects"
## [61] "has_downloads" "has_wiki"
## [63] "has_pages" "has_discussions"
## [65] "forks_count" "mirror_url"
## [67] "archived" "disabled"
## [69] "open_issues_count" "license"
## [71] "allow_forking" "is_template"
## [73] "web_commit_signoff_required" "topics"
## [75] "visibility" "forks"
```

```
## [77] "open_issues"          "watchers"
## [79] "default_branch"

names(jsonData$owner)

## [1] "login"          "id"              "node_id"
## [4] "avatar_url"     "gravatar_id"     "url"
## [7] "html_url"       "followers_url"   "following_url"
## [10] "gists_url"      "starred_url"     "subscriptions_url"
## [13] "organizations_url" "repos_url"       "events_url"
## [16] "received_events_url" "type"            "site_admin"

jsonData$owner$login

## [1] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
## [9] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
## [17] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
## [25] "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek" "jtleek"
```

1.8.3 Writing data frames to JSON

```
myjson <- toJSON(iris,pretty = TRUE)
cat(myjson)
```

```
## [
##   {
##     "Sepal.Length": 5.1,
##     "Sepal.Width": 3.5,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 4.9,
##     "Sepal.Width": 3,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 4.7,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 1.3,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 4.6,
##     "Sepal.Width": 3.1,
##     "Petal.Length": 1.5,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
##   },
##   {
##     "Sepal.Length": 5,
```

```

##      "Sepal.Width": 3.6,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.4,
##      "Sepal.Width": 3.9,
##      "Petal.Length": 1.7,
##      "Petal.Width": 0.4,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.6,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.3,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.4,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.9,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.1,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.4,
##      "Sepal.Width": 3.7,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.8,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 1.6,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##  },

```

```

## {
##   "Sepal.Length": 4.8,
##   "Sepal.Width": 3,
##   "Petal.Length": 1.4,
##   "Petal.Width": 0.1,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 4.3,
##   "Sepal.Width": 3,
##   "Petal.Length": 1.1,
##   "Petal.Width": 0.1,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.8,
##   "Sepal.Width": 4,
##   "Petal.Length": 1.2,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.7,
##   "Sepal.Width": 4.4,
##   "Petal.Length": 1.5,
##   "Petal.Width": 0.4,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.4,
##   "Sepal.Width": 3.9,
##   "Petal.Length": 1.3,
##   "Petal.Width": 0.4,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.1,
##   "Sepal.Width": 3.5,
##   "Petal.Length": 1.4,
##   "Petal.Width": 0.3,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.7,
##   "Sepal.Width": 3.8,
##   "Petal.Length": 1.7,
##   "Petal.Width": 0.3,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.1,
##   "Sepal.Width": 3.8,
##   "Petal.Length": 1.5,
##   "Petal.Width": 0.3,

```

```

##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.4,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 1.7,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.1,
##     "Sepal.Width": 3.7,
##     "Petal.Length": 1.5,
##     "Petal.Width": 0.4,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 4.6,
##     "Sepal.Width": 3.6,
##     "Petal.Length": 1,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.1,
##     "Sepal.Width": 3.3,
##     "Petal.Length": 1.7,
##     "Petal.Width": 0.5,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 4.8,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 1.9,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5,
##     "Sepal.Width": 3,
##     "Petal.Length": 1.6,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 1.6,
##     "Petal.Width": 0.4,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.2,
##     "Sepal.Width": 3.5,

```

```

##     "Petal.Length": 1.5,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.2,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 4.7,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 1.6,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 4.8,
##     "Sepal.Width": 3.1,
##     "Petal.Length": 1.6,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.4,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 1.5,
##     "Petal.Width": 0.4,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.2,
##     "Sepal.Width": 4.1,
##     "Petal.Length": 1.5,
##     "Petal.Width": 0.1,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 5.5,
##     "Sepal.Width": 4.2,
##     "Petal.Length": 1.4,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {
##     "Sepal.Length": 4.9,
##     "Sepal.Width": 3.1,
##     "Petal.Length": 1.5,
##     "Petal.Width": 0.2,
##     "Species": "setosa"
## },
## {

```

```

##      "Sepal.Length": 5,
##      "Sepal.Width": 3.2,
##      "Petal.Length": 1.2,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 3.5,
##      "Petal.Length": 1.3,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.9,
##      "Sepal.Width": 3.6,
##      "Petal.Length": 1.4,
##      "Petal.Width": 0.1,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.4,
##      "Sepal.Width": 3,
##      "Petal.Length": 1.3,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5.1,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 1.5,
##      "Petal.Width": 0.2,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 5,
##      "Sepal.Width": 3.5,
##      "Petal.Length": 1.3,
##      "Petal.Width": 0.3,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.5,
##      "Sepal.Width": 2.3,
##      "Petal.Length": 1.3,
##      "Petal.Width": 0.3,
##      "Species": "setosa"
##    },
##    {
##      "Sepal.Length": 4.4,
##      "Sepal.Width": 3.2,
##      "Petal.Length": 1.3,
##      "Petal.Width": 0.2,
##      "Species": "setosa"

```



```

## },
## {
##   "Sepal.Length": 5,
##   "Sepal.Width": 3.5,
##   "Petal.Length": 1.6,
##   "Petal.Width": 0.6,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.1,
##   "Sepal.Width": 3.8,
##   "Petal.Length": 1.9,
##   "Petal.Width": 0.4,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 4.8,
##   "Sepal.Width": 3,
##   "Petal.Length": 1.4,
##   "Petal.Width": 0.3,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.1,
##   "Sepal.Width": 3.8,
##   "Petal.Length": 1.6,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 4.6,
##   "Sepal.Width": 3.2,
##   "Petal.Length": 1.4,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5.3,
##   "Sepal.Width": 3.7,
##   "Petal.Length": 1.5,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 5,
##   "Sepal.Width": 3.3,
##   "Petal.Length": 1.4,
##   "Petal.Width": 0.2,
##   "Species": "setosa"
## },
## {
##   "Sepal.Length": 7,
##   "Sepal.Width": 3.2,
##   "Petal.Length": 4.7,

```

```

##     "Petal.Width": 1.4,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.4,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 4.5,
##     "Petal.Width": 1.5,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.9,
##     "Sepal.Width": 3.1,
##     "Petal.Length": 4.9,
##     "Petal.Width": 1.5,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5.5,
##     "Sepal.Width": 2.3,
##     "Petal.Length": 4,
##     "Petal.Width": 1.3,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.5,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 4.6,
##     "Petal.Width": 1.5,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5.7,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 4.5,
##     "Petal.Width": 1.3,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.3,
##     "Sepal.Width": 3.3,
##     "Petal.Length": 4.7,
##     "Petal.Width": 1.6,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 4.9,
##     "Sepal.Width": 2.4,
##     "Petal.Length": 3.3,
##     "Petal.Width": 1,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.6,

```

```

##      "Sepal.Width": 2.9,
##      "Petal.Length": 4.6,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.2,
##      "Sepal.Width": 2.7,
##      "Petal.Length": 3.9,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5,
##      "Sepal.Width": 2,
##      "Petal.Length": 3.5,
##      "Petal.Width": 1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.9,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.2,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6,
##      "Sepal.Width": 2.2,
##      "Petal.Length": 4,
##      "Petal.Width": 1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.1,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 4.7,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.6,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 3.6,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.7,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 4.4,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##  ],

```

```

## {
##   "Sepal.Length": 5.6,
##   "Sepal.Width": 3,
##   "Petal.Length": 4.5,
##   "Petal.Width": 1.5,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 5.8,
##   "Sepal.Width": 2.7,
##   "Petal.Length": 4.1,
##   "Petal.Width": 1,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 6.2,
##   "Sepal.Width": 2.2,
##   "Petal.Length": 4.5,
##   "Petal.Width": 1.5,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 5.6,
##   "Sepal.Width": 2.5,
##   "Petal.Length": 3.9,
##   "Petal.Width": 1.1,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 5.9,
##   "Sepal.Width": 3.2,
##   "Petal.Length": 4.8,
##   "Petal.Width": 1.8,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 6.1,
##   "Sepal.Width": 2.8,
##   "Petal.Length": 4,
##   "Petal.Width": 1.3,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 6.3,
##   "Sepal.Width": 2.5,
##   "Petal.Length": 4.9,
##   "Petal.Width": 1.5,
##   "Species": "versicolor"
## },
## {
##   "Sepal.Length": 6.1,
##   "Sepal.Width": 2.8,
##   "Petal.Length": 4.7,
##   "Petal.Width": 1.2,

```

```

##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.4,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 4.3,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.6,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.4,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.8,
##      "Sepal.Width": 2.8,
##      "Petal.Length": 4.8,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.7,
##      "Sepal.Width": 3,
##      "Petal.Length": 5,
##      "Petal.Width": 1.7,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 4.5,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.7,
##      "Sepal.Width": 2.6,
##      "Petal.Length": 3.5,
##      "Petal.Width": 1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 2.4,
##      "Petal.Length": 3.8,
##      "Petal.Width": 1.1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 2.4,

```

```

##      "Petal.Length": 3.7,
##      "Petal.Width": 1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.8,
##      "Sepal.Width": 2.7,
##      "Petal.Length": 3.9,
##      "Petal.Width": 1.2,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6,
##      "Sepal.Width": 2.7,
##      "Petal.Length": 5.1,
##      "Petal.Width": 1.6,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.4,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.5,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 4.5,
##      "Petal.Width": 1.6,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.7,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 4.7,
##      "Petal.Width": 1.5,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.3,
##      "Sepal.Width": 2.3,
##      "Petal.Length": 4.4,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.6,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.1,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {

```

```

##      "Sepal.Length": 5.5,
##      "Sepal.Width": 2.5,
##      "Petal.Length": 4,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.5,
##      "Sepal.Width": 2.6,
##      "Petal.Length": 4.4,
##      "Petal.Width": 1.2,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 6.1,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.6,
##      "Petal.Width": 1.4,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.8,
##      "Sepal.Width": 2.6,
##      "Petal.Length": 4,
##      "Petal.Width": 1.2,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5,
##      "Sepal.Width": 2.3,
##      "Petal.Length": 3.3,
##      "Petal.Width": 1,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.6,
##      "Sepal.Width": 2.7,
##      "Petal.Length": 4.2,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.7,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.2,
##      "Petal.Width": 1.2,
##      "Species": "versicolor"
##    },
##    {
##      "Sepal.Length": 5.7,
##      "Sepal.Width": 2.9,
##      "Petal.Length": 4.2,
##      "Petal.Width": 1.3,
##      "Species": "versicolor"

```

```

## },
## {
##     "Sepal.Length": 6.2,
##     "Sepal.Width": 2.9,
##     "Petal.Length": 4.3,
##     "Petal.Width": 1.3,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5.1,
##     "Sepal.Width": 2.5,
##     "Petal.Length": 3,
##     "Petal.Width": 1.1,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 5.7,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 4.1,
##     "Petal.Width": 1.3,
##     "Species": "versicolor"
## },
## {
##     "Sepal.Length": 6.3,
##     "Sepal.Width": 3.3,
##     "Petal.Length": 6,
##     "Petal.Width": 2.5,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 5.8,
##     "Sepal.Width": 2.7,
##     "Petal.Length": 5.1,
##     "Petal.Width": 1.9,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.1,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.9,
##     "Petal.Width": 2.1,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.3,
##     "Sepal.Width": 2.9,
##     "Petal.Length": 5.6,
##     "Petal.Width": 1.8,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.5,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.8,

```



```

##     "Petal.Width": 2.2,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.6,
##     "Sepal.Width": 3,
##     "Petal.Length": 6.6,
##     "Petal.Width": 2.1,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 4.9,
##     "Sepal.Width": 2.5,
##     "Petal.Length": 4.5,
##     "Petal.Width": 1.7,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.3,
##     "Sepal.Width": 2.9,
##     "Petal.Length": 6.3,
##     "Petal.Width": 1.8,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.7,
##     "Sepal.Width": 2.5,
##     "Petal.Length": 5.8,
##     "Petal.Width": 1.8,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.2,
##     "Sepal.Width": 3.6,
##     "Petal.Length": 6.1,
##     "Petal.Width": 2.5,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.5,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 5.1,
##     "Petal.Width": 2,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.4,
##     "Sepal.Width": 2.7,
##     "Petal.Length": 5.3,
##     "Petal.Width": 1.9,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.8,

```

```

##     "Sepal.Width": 3,
##     "Petal.Length": 5.5,
##     "Petal.Width": 2.1,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 5.7,
##     "Sepal.Width": 2.5,
##     "Petal.Length": 5,
##     "Petal.Width": 2,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 5.8,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 5.1,
##     "Petal.Width": 2.4,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.4,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 5.3,
##     "Petal.Width": 2.3,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.5,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.5,
##     "Petal.Width": 1.8,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.7,
##     "Sepal.Width": 3.8,
##     "Petal.Length": 6.7,
##     "Petal.Width": 2.2,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.7,
##     "Sepal.Width": 2.6,
##     "Petal.Length": 6.9,
##     "Petal.Width": 2.3,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6,
##     "Sepal.Width": 2.2,
##     "Petal.Length": 5,
##     "Petal.Width": 1.5,
##     "Species": "virginica"
## },

```

```

## {
##   "Sepal.Length": 6.9,
##   "Sepal.Width": 3.2,
##   "Petal.Length": 5.7,
##   "Petal.Width": 2.3,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 5.6,
##   "Sepal.Width": 2.8,
##   "Petal.Length": 4.9,
##   "Petal.Width": 2,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 7.7,
##   "Sepal.Width": 2.8,
##   "Petal.Length": 6.7,
##   "Petal.Width": 2,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.3,
##   "Sepal.Width": 2.7,
##   "Petal.Length": 4.9,
##   "Petal.Width": 1.8,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.7,
##   "Sepal.Width": 3.3,
##   "Petal.Length": 5.7,
##   "Petal.Width": 2.1,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 7.2,
##   "Sepal.Width": 3.2,
##   "Petal.Length": 6,
##   "Petal.Width": 1.8,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.2,
##   "Sepal.Width": 2.8,
##   "Petal.Length": 4.8,
##   "Petal.Width": 1.8,
##   "Species": "virginica"
## },
## {
##   "Sepal.Length": 6.1,
##   "Sepal.Width": 3,
##   "Petal.Length": 4.9,
##   "Petal.Width": 1.8,

```

```

##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.4,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 5.6,
##     "Petal.Width": 2.1,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.2,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.8,
##     "Petal.Width": 1.6,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.4,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 6.1,
##     "Petal.Width": 1.9,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.9,
##     "Sepal.Width": 3.8,
##     "Petal.Length": 6.4,
##     "Petal.Width": 2,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.4,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 5.6,
##     "Petal.Width": 2.2,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.3,
##     "Sepal.Width": 2.8,
##     "Petal.Length": 5.1,
##     "Petal.Width": 1.5,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 6.1,
##     "Sepal.Width": 2.6,
##     "Petal.Length": 5.6,
##     "Petal.Width": 1.4,
##     "Species": "virginica"
## },
## {
##     "Sepal.Length": 7.7,
##     "Sepal.Width": 3,

```

```

##      "Petal.Length": 6.1,
##      "Petal.Width": 2.3,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.3,
##      "Sepal.Width": 3.4,
##      "Petal.Length": 5.6,
##      "Petal.Width": 2.4,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.4,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 5.5,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6,
##      "Sepal.Width": 3,
##      "Petal.Length": 4.8,
##      "Petal.Width": 1.8,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.9,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 5.4,
##      "Petal.Width": 2.1,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.7,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 5.6,
##      "Petal.Width": 2.4,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 6.9,
##      "Sepal.Width": 3.1,
##      "Petal.Length": 5.1,
##      "Petal.Width": 2.3,
##      "Species": "virginica"
##    },
##    {
##      "Sepal.Length": 5.8,
##      "Sepal.Width": 2.7,
##      "Petal.Length": 5.1,
##      "Petal.Width": 1.9,
##      "Species": "virginica"
##    },
##    {

```

```

##     "Sepal.Length": 6.8,
##     "Sepal.Width": 3.2,
##     "Petal.Length": 5.9,
##     "Petal.Width": 2.3,
##     "Species": "virginica"
##   },
##   {
##     "Sepal.Length": 6.7,
##     "Sepal.Width": 3.3,
##     "Petal.Length": 5.7,
##     "Petal.Width": 2.5,
##     "Species": "virginica"
##   },
##   {
##     "Sepal.Length": 6.7,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.2,
##     "Petal.Width": 2.3,
##     "Species": "virginica"
##   },
##   {
##     "Sepal.Length": 6.3,
##     "Sepal.Width": 2.5,
##     "Petal.Length": 5,
##     "Petal.Width": 1.9,
##     "Species": "virginica"
##   },
##   {
##     "Sepal.Length": 6.5,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.2,
##     "Petal.Width": 2,
##     "Species": "virginica"
##   },
##   {
##     "Sepal.Length": 6.2,
##     "Sepal.Width": 3.4,
##     "Petal.Length": 5.4,
##     "Petal.Width": 2.3,
##     "Species": "virginica"
##   },
##   {
##     "Sepal.Length": 5.9,
##     "Sepal.Width": 3,
##     "Petal.Length": 5.1,
##     "Petal.Width": 1.8,
##     "Species": "virginica"
##   }
## ]

```

1.8.4 Convert back to JSON

```

iris.from <- fromJSON(myjson)
head(iris.from)

```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

1.8.5 Further resources

<http://www.json.org/> A good tutorial on jsonlite - <http://www.r-bloggers.com/new-package-jsonlite-a-smarter-json-encoderdecoder/> jsonlite vignette

1.9 The data.table Package

1.9.1 data.table

Inherits from data.frame - All functions that accept data.frame work on data.table Written in C so it is much faster Much, much faster at subsetting, group, and updating

1.9.2 Create data tables just like data frames

```
library(data.table)
DF <- data.frame(x = rnorm(9),
  y = rep(c("a", "b", "c"), each = 3),
  z = rnorm(9))
```

```
head(DF,3)
```

	x	y	z
	0.2112909	a	-0.6016176
	1.0683458	a	0.2184545
	-0.8828563	a	-1.1512846

```
DT = data.table(x=rnorm(9),y=rep(c("a", "b", "c"),each=3),z=rnorm(9))
head(DT,3)
```

	x	y	z
	-1.6202008	a	-0.1613972
	0.5406955	a	-0.8210609
	1.8305350	a	-0.1568646

1.9.3 See all the data tables in memory

```
tables()
```

```
##      NAME NROW NCOL MB  COLS KEY
## 1:   DT      9     3  0 x,y,z
## Total: 0MB
```

```
DT[2,]
```

x	y	z
0.5406955	a	-0.8210609

```
DT[DT$y=="a",]
```

x	y	z
-1.6202008	a	-0.1613972
0.5406955	a	-0.8210609
1.8305350	a	-0.1568646

1.9.4 Subsetting rows

```
DT[c(2,3)]
```

x	y	z
0.5406955	a	-0.8210609
1.8305350	a	-0.1568646

1.9.5 Subsetting columns!?

```
DT[,c(2,3)]
```

y	z
a	-0.1613972
a	-0.8210609
a	-0.1568646
b	2.2751699
b	0.8493389
b	0.7364842
c	-0.4208004
c	-0.1141261
c	0.9111860

1.9.6 Column subsetting in data.table

The subsetting function is modified for data.table The argument you pass after the comma is called an `l1expression` In R an expression is a collection of statements enclosed in curly brackets

```
{  
  X = 1  
  y = 2  
}  
k = {print(10); 5}
```

```
## [1] 10
```



```
print(k)
```

```
## [1] 5
```

1.9.7 Calculating values for variables with expressions

```
DT[,list(mean(x),sum(z))]
```

V1	V2
0.2343406	3.09793

```
DT[,table(y)]
```

```
## y
## a b c
## 3 3 3
```

1.9.8 Adding new columns

```
DT[,w:=z^2]
head(DT)
```

x	y	z	w
-1.6202008	a	-0.1613972	0.0260491
0.5406955	a	-0.8210609	0.6741409
1.8305350	a	-0.1568646	0.0246065
-0.4920395	b	2.2751699	5.1763980
0.1897013	b	0.8493389	0.7213766
0.9714955	b	0.7364842	0.5424090

```
DT2 <- DT
DT[,y:=2]
head(DT)
```

x	y	z	w
-1.6202008	2	-0.1613972	0.0260491
0.5406955	2	-0.8210609	0.6741409
1.8305350	2	-0.1568646	0.0246065
-0.4920395	2	2.2751699	5.1763980
0.1897013	2	0.8493389	0.7213766
0.9714955	2	0.7364842	0.5424090

```
head(DT,n=3)
```

x	y	z	w
-1.6202008	2	-0.1613972	0.0260491
0.5406955	2	-0.8210609	0.6741409
1.8305350	2	-0.1568646	0.0246065

```
head(DT2,n=3)
```

	x	y	z	w
	-1.6202008	2	-0.1613972	0.0260491
	0.5406955	2	-0.8210609	0.6741409
	1.8305350	2	-0.1568646	0.0246065

1.9.9 Multiple operations

```
DT[,m:= {tmp <- (x+z); log2(tmp+5)}]
head(DT)
```

	x	y	z	w	m
	-1.6202008	2	-0.1613972	0.0260491	1.686344
	0.5406955	2	-0.8210609	0.6741409	2.238675
	1.8305350	2	-0.1568646	0.0246065	2.738481
	-0.4920395	2	2.2751699	5.1763980	2.761951
	0.1897013	2	0.8493389	0.7213766	2.594319
	0.9714955	2	0.7364842	0.5424090	2.745878

1.9.10 plyr like operations

```
DT[,a:=x>0]
head(DT)
```

	x	y	z	w	m	a
	-1.6202008	2	-0.1613972	0.0260491	1.686344	FALSE
	0.5406955	2	-0.8210609	0.6741409	2.238675	TRUE
	1.8305350	2	-0.1568646	0.0246065	2.738481	TRUE
	-0.4920395	2	2.2751699	5.1763980	2.761951	FALSE
	0.1897013	2	0.8493389	0.7213766	2.594319	TRUE
	0.9714955	2	0.7364842	0.5424090	2.745878	TRUE

```
DT[,b:= mean(x+w) ,by=a]
head(DT)
```

	x	y	z	w	m	a	b
	-1.6202008	2	-0.1613972	0.0260491	1.686344	FALSE	0.6080906
	0.5406955	2	-0.8210609	0.6741409	2.238675	TRUE	1.4116886
	1.8305350	2	-0.1568646	0.0246065	2.738481	TRUE	1.4116886
	-0.4920395	2	2.2751699	5.1763980	2.761951	FALSE	0.6080906
	0.1897013	2	0.8493389	0.7213766	2.594319	TRUE	1.4116886
	0.9714955	2	0.7364842	0.5424090	2.745878	TRUE	1.4116886

1.9.11 Special variables

N An integer, length 1, containing the number

```
set.seed(123)
DT <- data.table(x=sample(letters[1:3],1E5,TRUE))
head(DT)
```

```

  x
  c
  c
  c
  b
  c
  b

```

```
DT[, .N, by=x]
```

x	N
c	33294
b	33305
a	33401

1.9.12 Keys

```
DT <- data.table(x=rep(c("a","b","c"),each=100), y=rnorm(300))
head(DT)
```

x	y
a	0.8863126
a	2.8285813
a	2.0314543
a	1.9067541
a	0.2149083
a	-0.8627341

```
setkey(DT,x)
```

```
DT["a"]
```

x	y
a	0.8863126
a	2.8285813
a	2.0314543
a	1.9067541
a	0.2149083
a	-0.8627341
a	-2.2049386
a	0.2410592
a	1.8383242
a	0.7920547
a	0.6505347

x	y
a	-1.5391206
a	-0.6083005
a	0.3819564
a	-1.0750004
a	0.2199426
a	-0.7828878
a	-1.1100335
a	-1.6587146
a	-0.5014734
a	1.9163637
a	1.4123664
a	0.9226099
a	1.0110620
a	0.5721303
a	-0.6284313
a	-0.3631614
a	-1.0585881
a	-0.4293580
a	0.8694147
a	-0.5400165
a	-1.1464775
a	-0.1715184
a	-0.5636834
a	-0.4299435
a	-1.2372378
a	0.1590133
a	-1.1671107
a	-0.0811194
a	-0.5166795
a	0.9954070
a	0.7975214
a	0.5389522
a	-1.4040561
a	0.4014406
a	-0.5243224
a	-0.8395215
a	0.4755659
a	-0.0119470
a	0.1031978
a	-0.3857541
a	1.1172644
a	-0.4996139
a	-0.4473509
a	-0.2378451
a	-0.8693937
a	1.1488768
a	0.5386500
a	-0.1068099
a	0.6005365
a	-1.4749944
a	0.9812696
a	-0.6111874

x	y
a	0.0893865
a	-0.0132723
a	-0.9721934
a	-0.5794623
a	0.1496314
a	0.4764069
a	0.4472968
a	-0.1918096
a	0.5171271
a	0.4033827
a	1.7841138
a	0.2777564
a	0.7739498
a	-2.0808193
a	-0.3592089
a	-0.4593222
a	0.2018195
a	0.6240114
a	-0.2572298
a	0.9441402
a	0.2507481
a	-0.7278426
a	0.3688132
a	0.4441507
a	-1.0053542
a	-0.3315247
a	-0.3703933
a	-0.7970153
a	0.2814856
a	0.3330725
a	0.5269032
a	-0.7816895
a	-0.0279395
a	-1.7449234
a	0.6528421
a	-0.9383082
a	0.6275316

1.9.13 Joins

```
DT1 <- data.table(x=c('a', 'a', 'b', 'dt1'), y=1:4)
DT2 <- data.table(x=c('a', 'b', 'dt2'), z=5:7)
setkey(DT1, x); setkey(DT2, x)
merge(DT1, DT2)
```

x	y	z
a	1	5
a	2	5
b	3	6

1.9.14 Fast reading

```
big_df <- data.frame(x=rnorm(1E6), y=rnorm(1E6))
file<- tempfile()
write.table(big_df, file=file, row.names=FALSE, col.names=TRUE, sep="\t", quote=FALSE)
system.time(fread(file))
```

```
##      user  system elapsed
##    0.042    0.004    0.046
```

```
system.time(read.table(file, header=TRUE, sep="\t"))
```

```
##      user  system elapsed
##    2.330    0.033    2.363
```

1.9.15 Summary and further reading

The latest development version contains new functions like melt and dcast for data.tables - <https://r-forge.r-project.org/scm/viewvc.php/pkg/NEWS?view=markup&root=datatable> Here is a list of differences between data.table and data.frame - <http://stackoverflow.com/questions/13618488/what-features-can-you-do-with-data-frame-that-you-cant-do-with-data-table> Notes based on Raphael Gottardo's notes https://github.com/raphg/Bios@578/blob/master/Advanced_data_manipulation.Rpres, who got them from Kevin Ushey.