

# Data Science Course Notes

*Joao Pinelo Silva*

*September 11, 2015*

## Calculating / Estimating Memory Requirements

DATA: Data frame with 1,500,000 rows and 120 columns of numeric data and a 64 bits

Formula

*Number of rows x Number of columns x 8 bytes*

Why 8 bytes? machine/OS is 64 bits = 8 bytes, since there are 8 bits per byte

$1,500,000 \times 120 \times 8 = 1440000000$

$1440000000 / 2^{20} \text{ bytes/MB} = 1,373.29 \text{ MB} = 1.34 \text{ GB}$

## R Connections

For example opening, reading and or writing files, open URL...

## Create File

```
dput(x, file = "dputed_x.R")
```

file() basic arguments:

- description = name of file;
- open = code indicating:
  - “r” read only,
  - “w” writing and initializing a new file,
  - “a” appending,
  - “rb”, “wb”, “ab” reading, writing, or appending in binary mode (windows)

## Subsetting

- [ returns object of same class as original - can be used to extract more than one element (one exception)
- [[ extract elements of list or df. Can extract a single element. Not necessarily return list or df.
- \$ extract elements of list or df by name