1. Week 1

1.1 Motivation & Pre-requisites

What does data really look like?

<http://brianknaus.com/software/srtoolbox/s_4_1_sequence80.txt>

<https://dev.twitter.com/docs/api/1/get/blocks/blocking>

<http://blue-button.github.com/challenge/>

Where is data?

<http://rickosborne.org/blog/2010/02/infographic-migrating-from-sql-to-mapreduce-with-mongodb/>

<https://dev.twitter.com/docs/api/1/get/blocks/blocking>

<https://data.baltimorecity.gov/>

1.2. Raw & Processed Data

Definition of Data <http://en.wikipedia.org/wiki/Data>

Raw versus Processed Data <http://en.wikipedia.org/wiki/Raw_data>

An Example of a Processing Pipeline <http://www.illumina.com.cn/support/sequencing/sequencing_instruments/hiseq_1000.asp>

<http://www.cbcb.umd.edu/~hcorrada/CMSC858B/lectures/lect22_seqIntro/seqIntro.pdf>

1.3. The Components of Tidy Data

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

Why is the Instruction List Important <http://www.colbertnation.com/the-colbert-report-videos/425748/april-23-2013/austerity-s-spreadsheet-error>

1.4. Downloading Files &

1.5. Reading Local Flat Files

Example - Baltimore camera data <https://data.baltimorecity.gov/Transportation/Baltimore-Fixed-Speed-Cameras/dz54-2aru>

1.6. Reading Excel Files

Excel Files <http://office.microsoft.com/en-us/excel/>

Example - Baltimore camera data <https://data.baltimorecity.gov/Transportation/Baltimore-Fixed-Speed-Cameras/dz54-2aru>

Further Notes

<http://cran.r-project.org/web/packages/XLConnect/index.html>

<http://cran.r-project.org/web/packages/XLConnect/vignettes/XLConnect.pdf>

1.7. Reading XML

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

XPath <http://www.stat.berkeley.edu/~statcur/Workshop2/Presentations/XML.pdf>

Another example <http://espn.go.com/nfl/team/_/name/bal/baltimore-ravens>

Notes & further resources

<http://www.omegahat.org/RSXML/shortIntro.pdf>

<http://www.omegahat.org/RSXML/Tour.pdf>

<http://www.stat.berkeley.edu/~statcur/Workshop2/Presentations/XML.pdf>

1.8. Reading JSON

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

Convert back to JSON <http://www.r-bloggers.com/new-package-jsonlite-a-smarter-json-encoderdecoder/>

Further Resources

<http://www.json.org/>

<http://www.r-bloggers.com/new-package-jsonlite-a-smarter-json-encoderdecoder/>

<http://cran.r-project.org/web/packages/jsonlite/vignettes/json-mapping.pdf>

1.9. Using data.table

Summary & further reading

<https://r-forge.r-project.org/scm/viewvc.php/pkg/NEWS?view=markup&root=datatablehttp://stackoverflow.com/questions/13618488/what-you-can-do-with-data-frame-that-you-cant-in-data-table>

<http://stackoverflow.com/questions/13618488/what-you-can-do-with-data-frame-that-you-cant-in-data-table>

<https://github.com/raphg/Biostat-578/blob/master/Advanced_data_manipulation.Rpres>