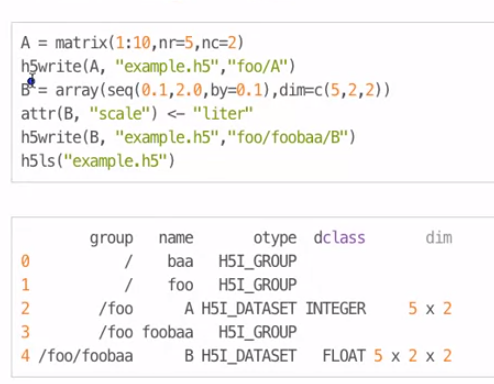
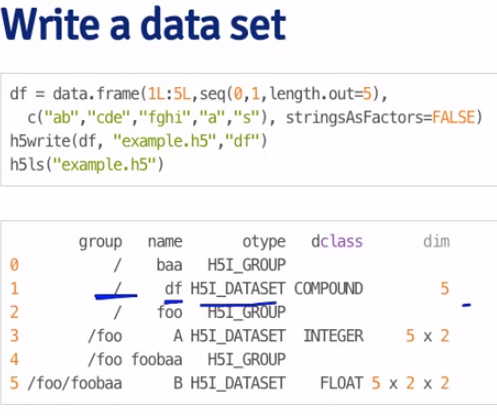
22 Getting Data Video Lecture – Reading from HDF5

**HDF5 Overview**

1. Used for storing large dataset
2. Used for storing structured dataset
3. HDF = hierarchical data format
4. data is stored in *groups* containing zero or more datasets and metadata
   1. Have a *group header* with group name and a list of attributes
   2. Have a *group symbol table* with a list of objects in group
5. *datasets* are multidimensional arrays of data elements with metadata
   1. they have a *header* with the name, datatype, dataspace, and storage layout
   2. they have a *data array* with the data like a data frame

**Creating a RHDF5 data set in R**

1. library(rhdf5)
2. created = h5createFile("example.h5")
3. check out <http://www.bioconductor.org/packages//2.13/bioc/vignettes/rhdf5/inst/doc/rhdf5.pdf> for more on using R with Rhdf5
4. create groups within the file:
5. created = h5createGroup(“example.h5”,”foo”) \*\*creates a group within file example.h5 named “foo”
6. created = h5createGroup(“example.h5”, “foo/foobaa”) creates a subgroup within foo named foobaa that is within the file example.h5
7. h5ls(“example.h5”) \*\*shows the structure of the h5 file example.h5
8. to fill in the groups of your h5 file:
   1. 
   2. 
9. Reading H5 files
   1. readA = h5read(“example.h5”, “foo/A”)
   2. readdf = h5read(example.h5, “df”)
   3. writing to a chunk of a dataset of an h5 file
      1. 
      2. where the index is given at write to the first 3 rows in the first column of “foo/A”
      3. the indexing command can also be given to h5read