Distance Metrics

Mark Voorhies

5/22/2013

Expression profiling pipelines

RnaSeq pipelines

Fun with logarithms

In log space, multiplication and division become addition and subtraction:

$$\begin{aligned} \log(xy) &= \log(x) + \log(y) \\ \log(x/y) &= \log(x) - \log(y) \end{aligned}$$

Fun with logarithms

In log space, multiplication and division become addition and subtraction:

$$\begin{aligned} \log(xy) &= \log(x) + \log(y) \\ \log(x/y) &= \log(x) - \log(y) \end{aligned}$$

Therefore, exponentiation becomes multiplication:

$$\log(x^y) = y \log(x)$$

Fun with logarithms

In log space, multiplication and division become addition and subtraction:

$$\begin{aligned} \log(xy) &= \log(x) + \log(y) \\ \log(x/y) &= \log(x) - \log(y) \end{aligned}$$

Therefore, exponentiation becomes multiplication:

$$\log(x^y) = y \log(x)$$

Also, we can change of the base of a logarithm like so:

$$\log_A(x) = \log(x)/\log(A)$$

Review

- Watch out for spelling errors *(tab-completion can help)*

Review

- Watch out for spelling errors *(tab-completion can help)*
- Use \." to \look inside" objects

Review

- Watch out for spelling errors *(tab-completion can help)*
- Use \." to \look inside" objects
- Think about appropriate inputs and outputs for your functions

List tricks

Adding data to a list:

```
mylist = []
mylist.append(3)
mylist += [4,5,6]
```

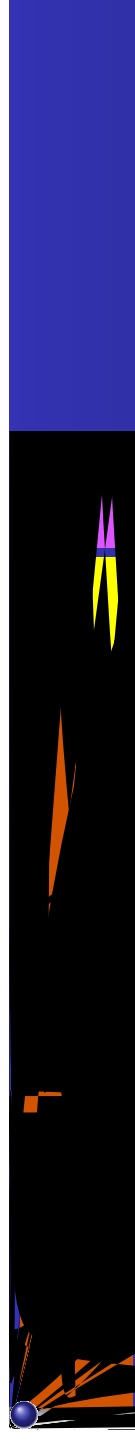
List tricks

Adding data to a list:

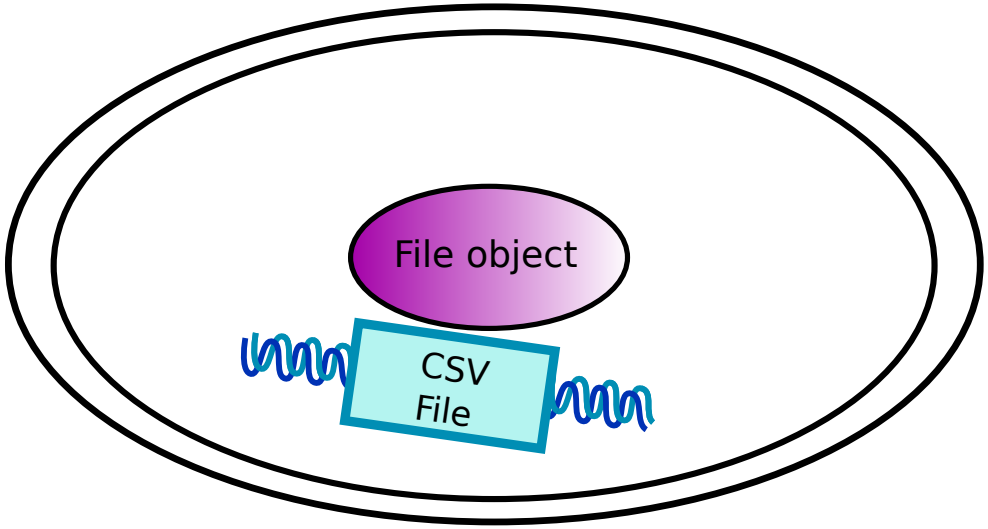
```
mylist = []
mylist.append(3)
mylist += [4,5,6]
```

Lists of lists:

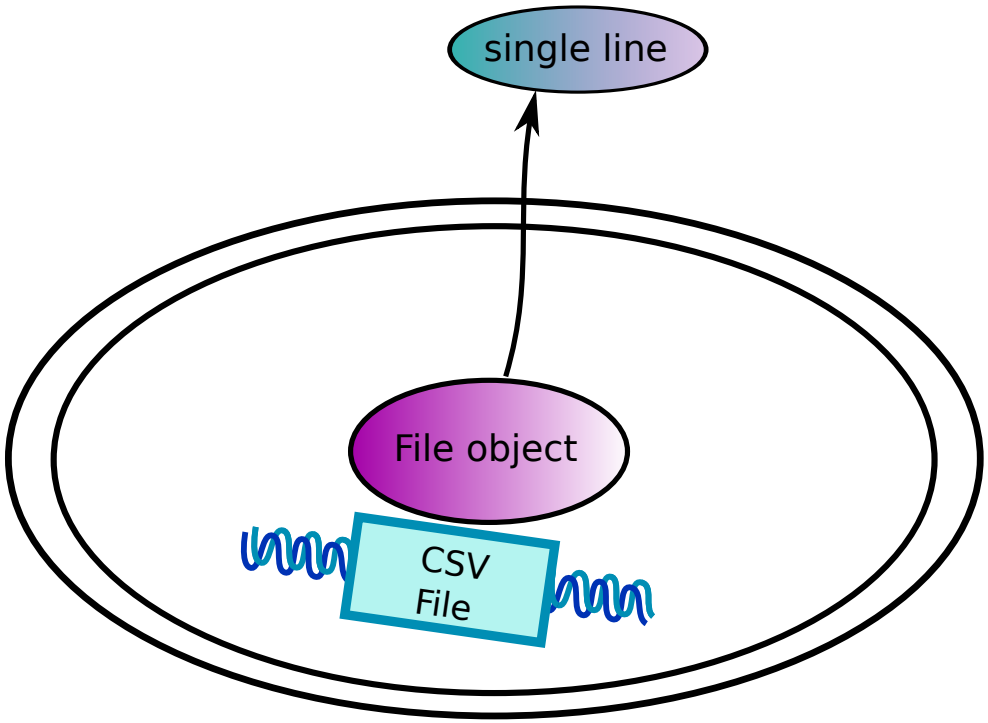
```
matrix = [[ 1, 2, 3, 4],
          [ 5, 6, 7, 8],
          [ 9,10,11,12]]
```



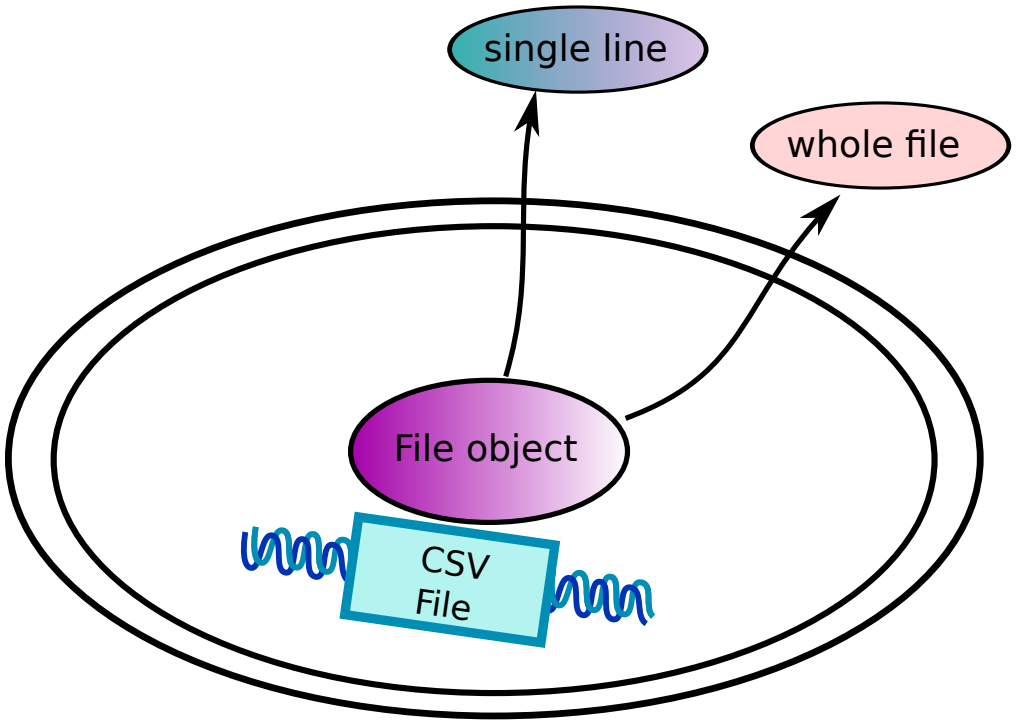
open(“supp2data.csv”)



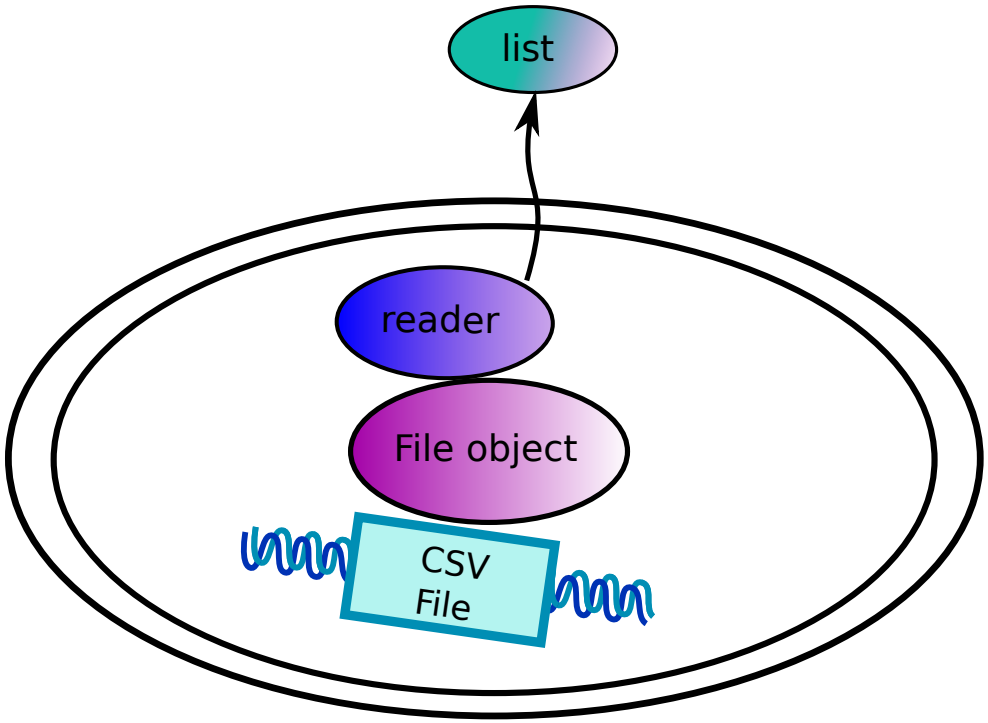
open(“supp2data.csv”).next()



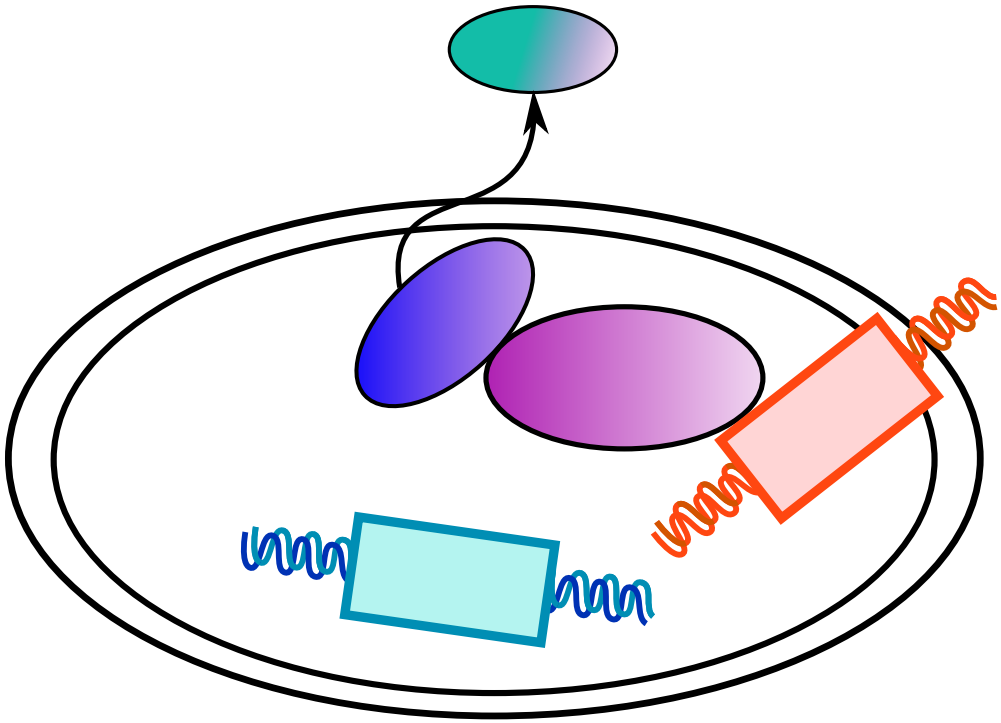
open("supp2data.csv").read()



csv.reader(open("supp2data.csv")).next()



csv.reader(urlopen("http://example.com/csv")).next()



Homework

1. Write a function to calculate all pairwise Pearson correlations for the yeast expression pro les.
2. Save the results of your pairwise correlation calculation in the CDT format described in the JavaTreeView manual.
3. Read PNAS 95:14863
4. Try the rst two problems, replacing the Pearson correlation with the distance metric from the PNAS paper or with one of the distance metrics from the Cluster3 manual.