Pandas

Slice rows

Df.ix[‘two’:’four’]

You can get an intersection of a row and column

Df.ix[[‘two’, ‘three’], [‘A’, ‘B’]

Start, Stop and Step,

You can have a step value of 3, and just take 4-7 in columns

How does pandas handle missing data?

Take two series, stick them together, and we get one dataframe WITH MISSING VALUES

S1 =pd.Series([1,2,3,4], index=list(‘ABCD’)

S2=pd.Series[5,6,7,8], index=list[c,d,e,f]

When we have missing values, we can pass default of -9999 or 0, or you can ERASE them

Df.dropna()

Df.fillna(0)

Df.reindex(index=…) fill in data for remaining cells we don’t have. We can fillin with value etc.

Indices

* Dataframes are built along an index
* Standard index, multiindex
* Int64index, datetimeindex, periodindex
  + Based on timeseries, or ranges of things, quarterlies etc.

MultiIndex / Hierarchial index

Datafarme with a hierarchial index → n dimention array

e.g. different segments of customers, days of the week, and different tests

multiindex is a multiple of tuples

Creating a multiIndex

* 4 separate items, primary keys in database
* you can get tuples, or a constructor
* two different lists etc.

Merging

* database style joins
* merging, joining concatenating
* if you can have one data frame, you can merge them
* joined based on a same column name
* mergin on different columns
  + you can join on a left column, right column, and it aligns them so it matches up
* if you have a nindex, you can join on index
* left\_index=true, right\_on=’y’

Join with a missing A value