Algorithmic Approaches for Biological Data, Lecture #12

Katherine St. John

City University of New York
American Museum of Natural History

2 March 2016

Outline

- File Overview:

Outline

- File Overview:
  - Standard I/O

Outline

- File Overview:
  - Standard I/O
  - CSV reader

Outline

- File Overview:
  - Standard I/O
  - CSV reader
  - matplotlib's imread()

Outline

- File Overview:
  - Standard I/O
  - CSV reader
  - matplotlib's imread()
  - The `urllib` reader

Outline

- File Overview:
  - Standard I/O
  - CSV reader
  - matplotlib's imread()
  - The `urllib` reader
- Structured Data files: CSV, PNG (Image), SQL, FASTA & friends

Standard File Input/Output

infile.txt
Hello! This is a test. 123

Standard File Input/Output

infile.txt
Hello! This is a test. 123

```
"Hello!\nThis is \na \ntest.\n123"
```

Standard File Input/Output

infile.txt
Hello! This is a test. 123

```
"Hello!\nThis is \na \ntest.\n123"
```

- Text files are multi-lined strings.

Standard File Input/Output

infile.txt
Hello! This is a test. 123

```
"Hello!\nThis is \na \ntest.\n123"
```

- Text files are multi-lined strings.
- Lines are indicated by '\n' characters.

Files Commands

- Opening a file:

Files Commands

- Opening a file:
  `infile = open('data.txt', 'r')`

Files Commands

- Opening a file:

```
infile = open('data.txt', 'r')
outfile = open('log.txt', 'w')
```

Files Commands

- Opening a file:

```
infile = open('data.txt', 'r')
outfile = open('log.txt', 'w')
```

- Reading from a file:

Files Commands

- Opening a file:
  ```
  infile = open('data.txt', 'r')
  outfile = open('log.txt', 'w')
  ```
- Reading from a file:
  - `infile.read()`: reads the entire file into a single string.

Files Commands

- Opening a file:
  ```
  infile = open('data.txt', 'r')
  outfile = open('log.txt', 'w')
  ```
- Reading from a file:
  - `infile.read()`: reads the entire file into a single string.
  - `infile.readline()`: read the next line of the file.

Files Commands

- Opening a file:
  ```
  infile = open('data.txt', 'r')
  outfile = open('log.txt', 'w')
  ```
- Reading from a file:
  - `infile.read()`: reads the entire file into a single string.
  - `infile.readline()`: read the next line of the file.
  - `infile.readlines()`: read the file into a list of strings.

Files Commands

- Opening a file:
```
infile = open('data.txt', 'r')
outfile = open('log.txt', 'w')
```
- Reading from a file:
  - `infile.read()`: reads the entire file into a single string.
  - `infile.readline()`: read the next line of the file.
  - `infile.readlines()`: read the file into a list of strings.
- Closing a file:

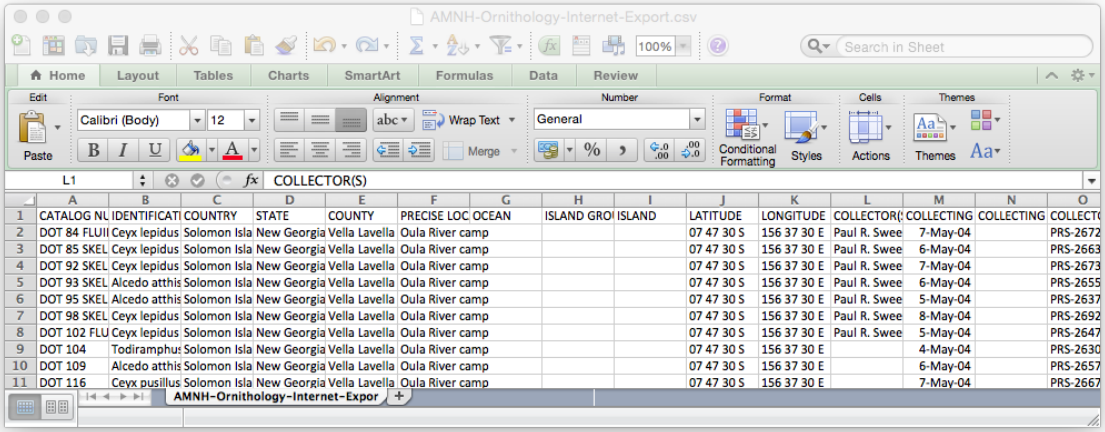
Files Commands

- Opening a file:
  `infile = open('data.txt', 'r')`
  `outfile = open('log.txt', 'w')`
- Reading from a file:
  - `infile.read()`: reads the entire file into a single string.
  - `infile.readline()`: read the next line of the file.
  - `infile.readlines()`: read the file into a list of strings.
- Closing a file:
  `infile.close()`
- Writing to a file:

Files Commands

- Opening a file:
  `infile = open('data.txt', 'r')`
  `outfile = open('log.txt', 'w')`
- Reading from a file:
  - `infile.read()`: reads the entire file into a single string.
  - `infile.readline()`: read the next line of the file.
  - `infile.readlines()`: read the file into a list of strings.
- Closing a file:
  `infile.close()`
- Writing to a file:
  `outfile.write(s)`

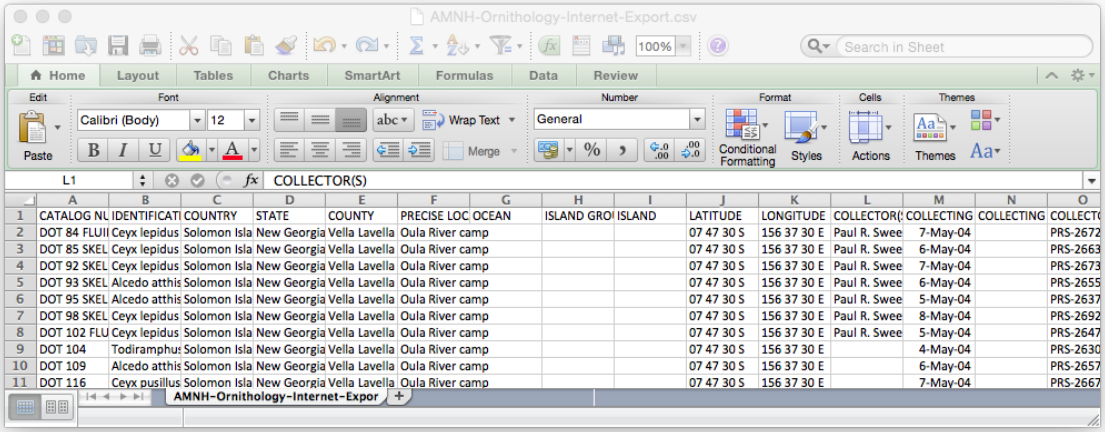
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(S	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUII	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Very structured– the columns and rows matter.

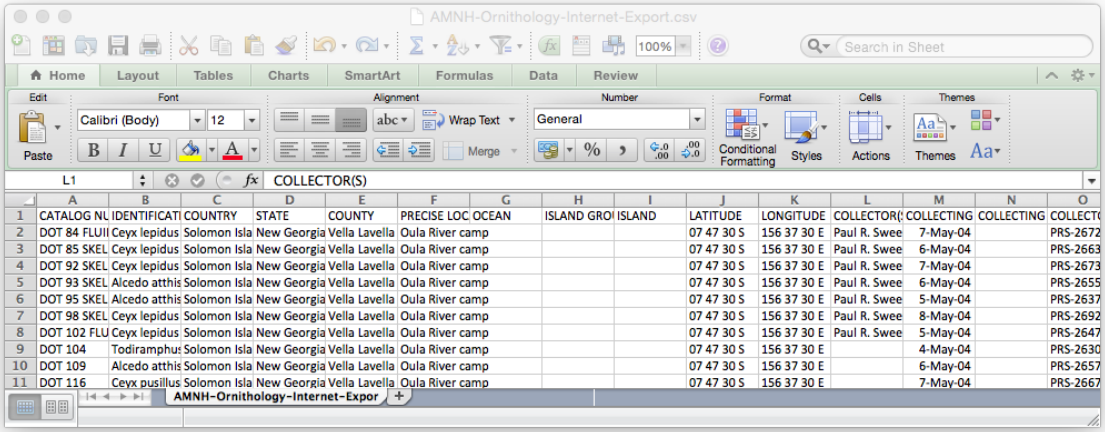
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(S	COLLECTING	COLLECTING	COLLECTO
2	DOT 84 FLUII	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Very structured– the columns and rows matter.
- To keep that format as a text file:

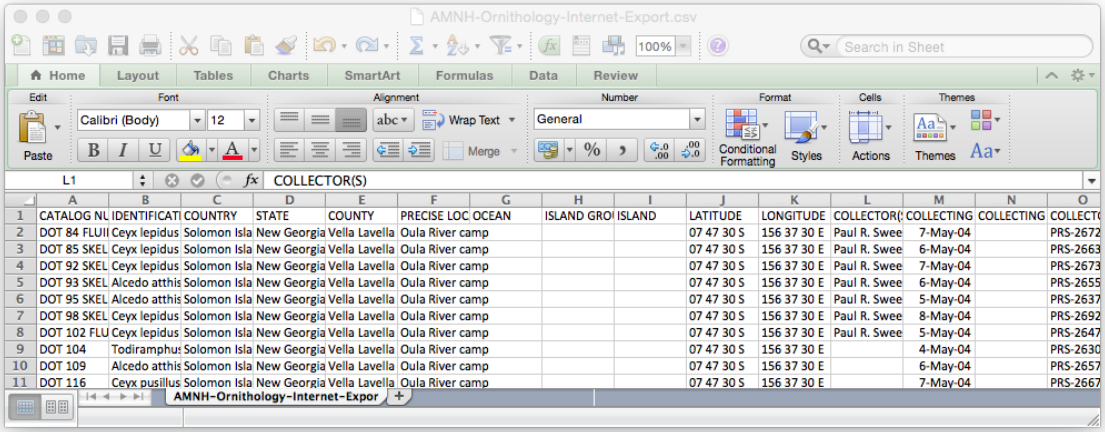
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(S	COLLECTING	COLLECTING	COLLECTO
2	DOT 84 FLUII	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Very structured– the columns and rows matter.
- To keep that format as a text file:
  - columns separated by commas (',') and

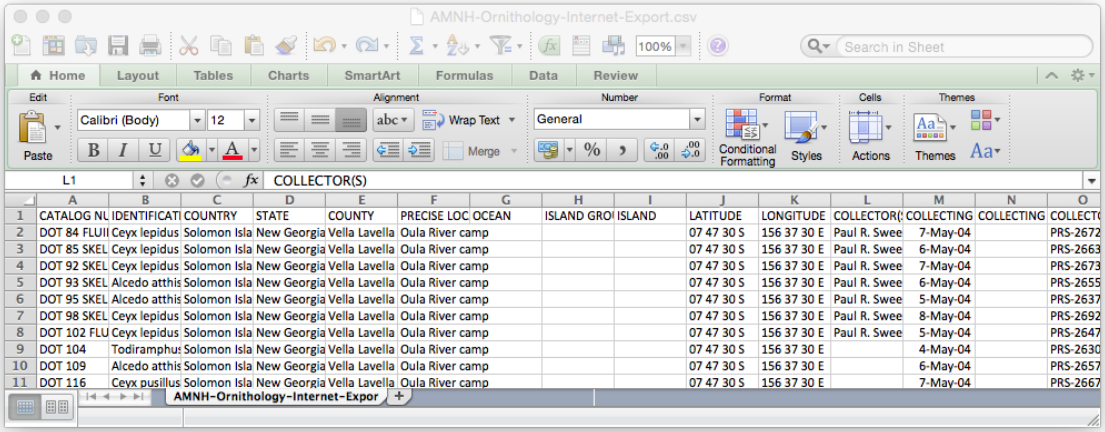
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUI	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Very structured– the columns and rows matter.
- To keep that format as a text file:
  - columns separated by commas (',') and
  - rows separated by new lines ('\n')

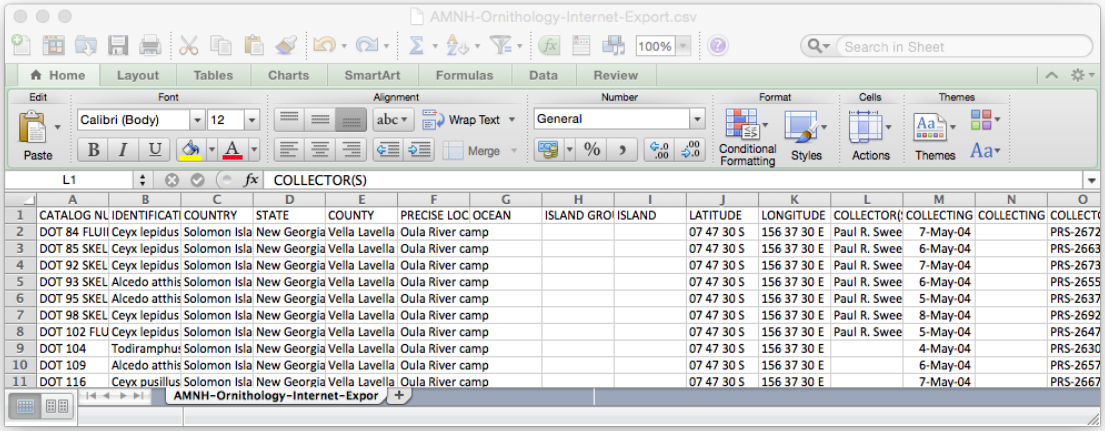
CSV Files



- Very structured– the columns and rows matter.
- To keep that format as a text file:
  - columns separated by commas (',') and
  - rows separated by new lines ('\n')
- Rows look like:

```
"DOT 84 FLUID 11383",Ceyx lepidus collectoris,Solomon Islands,New Georgia Group,Vella Lavella
Island,Oula River camp,,,,07 47 30 S,156 37 30 E,Paul R. Sweet,7-May-04,,PRS-2672,,,"Tissue Fluid "
```

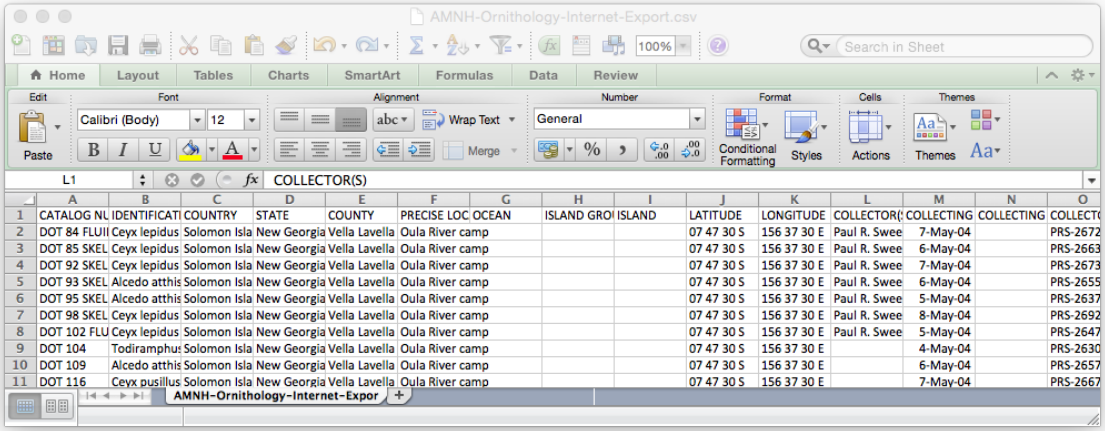
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUI	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Built-in package for reading CSV files. To use it:

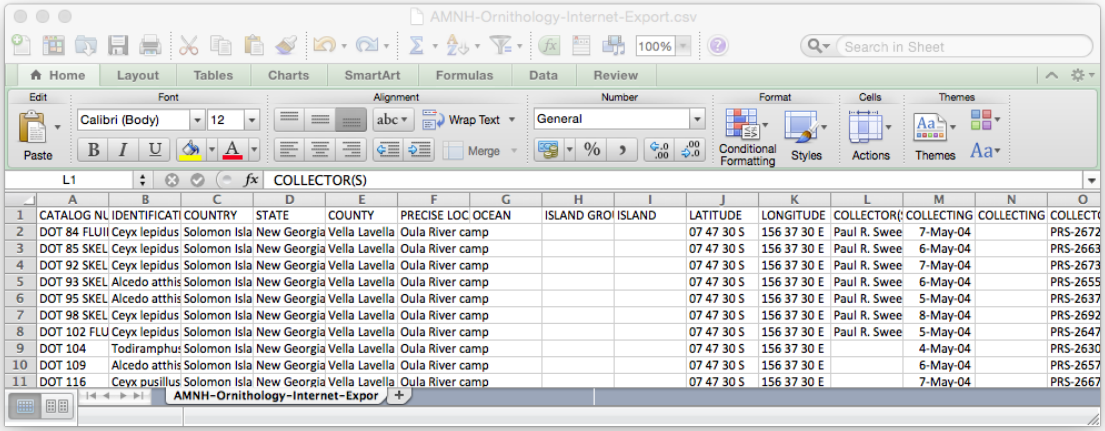
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUI	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`

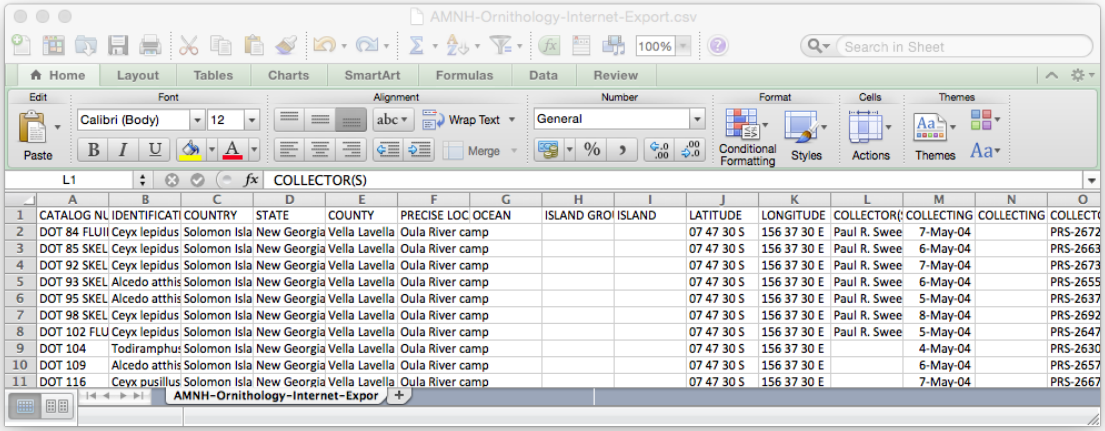
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUI	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`
  - Open the file normally: `f = open("in.csv", "rU")`

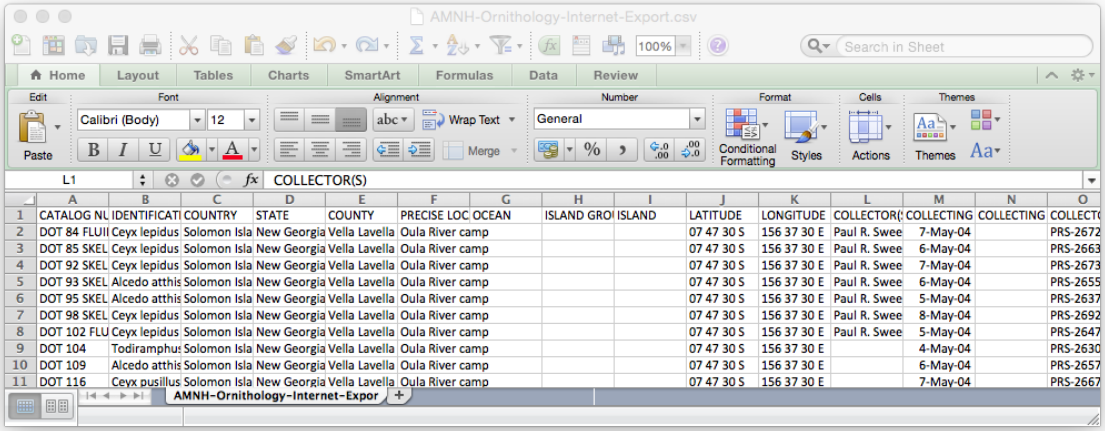
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUI	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

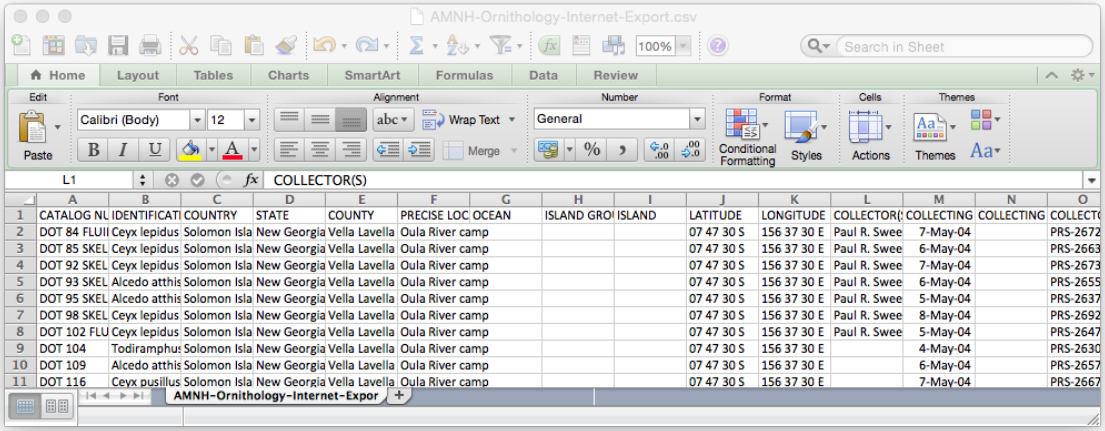
- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`
  - Open the file normally: `f = open("in.csv", "rU")`
    `"rU"` avoids errors with different newlines, and accepts all variants.

CSV Files



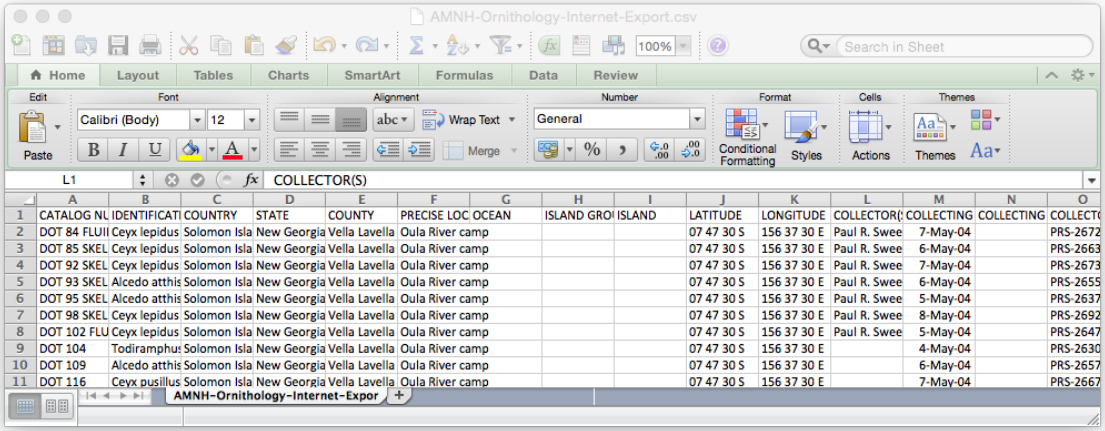
- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`
  - Open the file normally: `f = open("in.csv", "rU")`
    `"rU"` avoids errors with different newlines, and accepts all variants.
  - Create a reader: `reader = csv.DictReader(f)`

CSV Files



- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`
  - Open the file normally: `f = open("in.csv", "rU")`
    `"rU"` avoids errors with different newlines, and accepts all variants.
  - Create a reader: `reader = csv.DictReader(f)`
    Uses column names in first line of csv file to access row data.

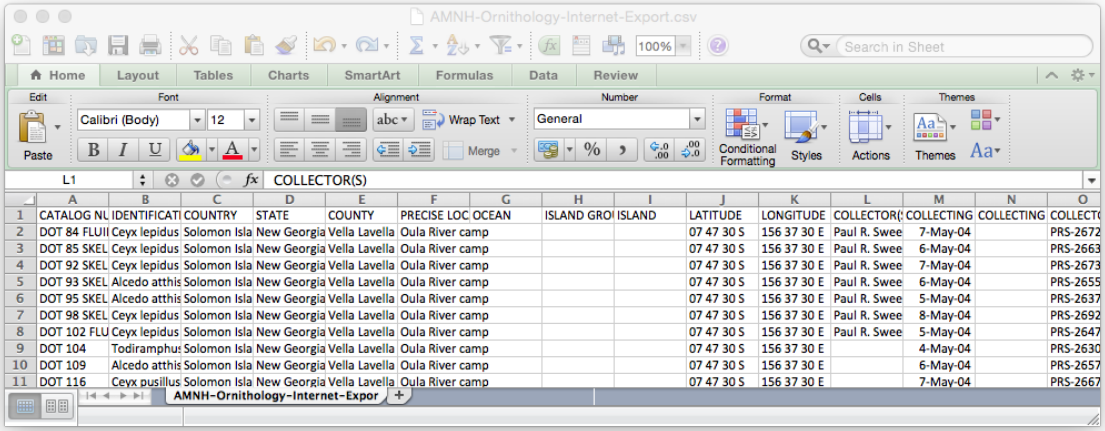
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUI	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`
  - Open the file normally: `f = open("in.csv", "rU")`
    `"rU"` avoids errors with different newlines, and accepts all variants.
  - Create a reader: `reader = csv.DictReader(f)`
    Uses column names in first line of csv file to access row data.
  - Read in lines from reader:

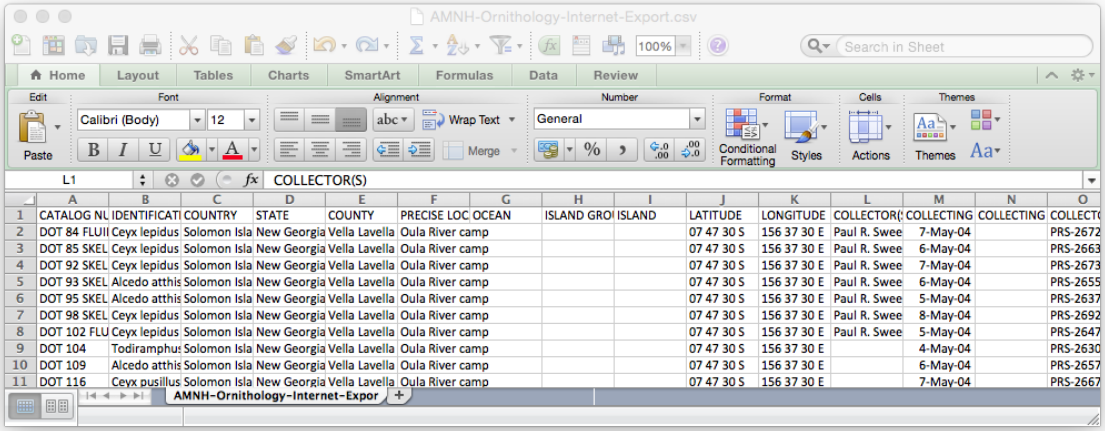
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUI	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`
  - Open the file normally: `f = open("in.csv", "rU")`
    `"rU"` avoids errors with different newlines, and accepts all variants.
  - Create a reader: `reader = csv.DictReader(f)`
    Uses column names in first line of csv file to access row data.
  - Read in lines from reader: `for row in reader:`

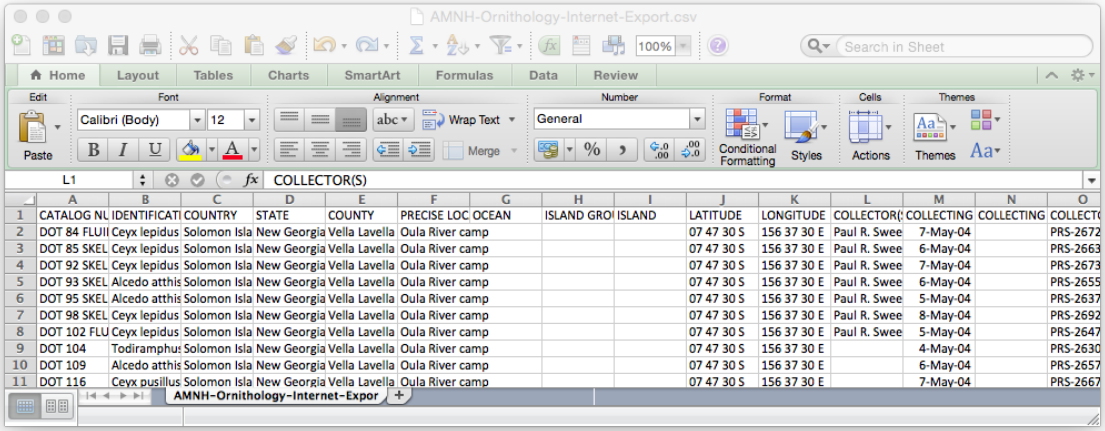
CSV Files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CATALOG NU	IDENTIFICATI	COUNTRY	STATE	COUNTY	PRECISE LOC	OCEAN	ISLAND GRO	ISLAND	LATITUDE	LONGITUDE	COLLECTOR(	COLLECTING	COLLECTING	COLLECT
2	DOT 84 FLUII	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2672
3	DOT 85 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2663
4	DOT 92 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	7-May-04		PRS-2673
5	DOT 93 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	6-May-04		PRS-2655
6	DOT 95 SKEL	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2637
7	DOT 98 SKEL	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	8-May-04		PRS-2692
8	DOT 102 FLU	Ceyx lepidus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E	Paul R. Swee	5-May-04		PRS-2647
9	DOT 104	Todiramphus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		4-May-04		PRS-2630
10	DOT 109	Alcedo atthis	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		6-May-04		PRS-2657
11	DOT 116	Ceyx pusillus	Solomon Isla	New Georgia	Vella Lavella	Oula River camp				07 47 30 S	156 37 30 E		7-May-04		PRS-2667

- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`
  - Open the file normally: `f = open("in.csv", "rU")`
    `"rU"` avoids errors with different newlines, and accepts all variants.
  - Create a reader: `reader = csv.DictReader(f)`
    Uses column names in first line of csv file to access row data.
  - Read in lines from reader: `for row in reader:`
  - To access individual entries in a row:

CSV Files



- Built-in package for reading CSV files. To use it:
  - At top of file, include: `import csv`
  - Open the file normally: `f = open("in.csv", "rU")`
    `"rU"` avoids errors with different newlines, and accepts all variants.
  - Create a reader: `reader = csv.DictReader(f)`
    Uses column names in first line of csv file to access row data.
  - Read in lines from reader: `for row in reader:`
  - To access individual entries in a row:
    `if "Malaysia" in row['COUNTRY']: ...`

Structured Data Files

- CSV: comma separated values

Structured Data Files

- CSV: comma separated values
- FASTA: sequence data

```
> Rosalind.1289
ACGTTTAATTATATATAAAA
> Rosalind.1988
GCGAAGGTATTTGGSAAAA

GCGCGCGCGCCCCTTATAT
```

Structured Data Files

- CSV: comma separated values
- FASTA: sequence data

```
> Rosalind.1289
ACGTTTAATTATATATAAAA
> Rosalind.1988
GCGAAGGTATTTGGSAAAA

GCGCGCGCGCCCCTTATAT
```

- Portable Network Graphics (PNG): stores images pixel by pixel.

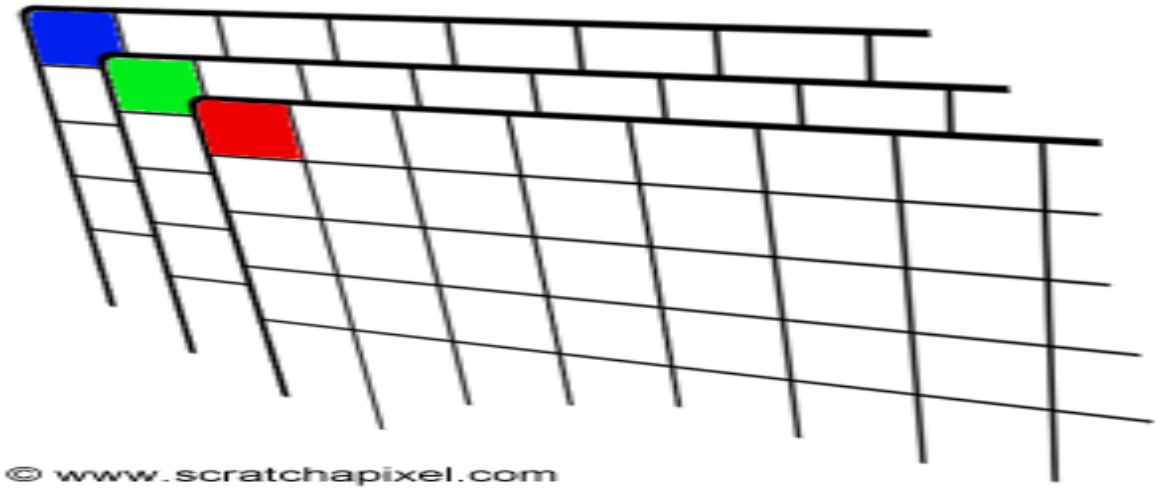
Structured Data Files

- CSV: comma separated values
- FASTA: sequence data

```
> Rosalind.1289
ACGTTTAATTATATATAAAA
> Rosalind.1988
GCGAAGGTATTTGGSAAAA

GCGCGCGCGCCCCTTATAT
```

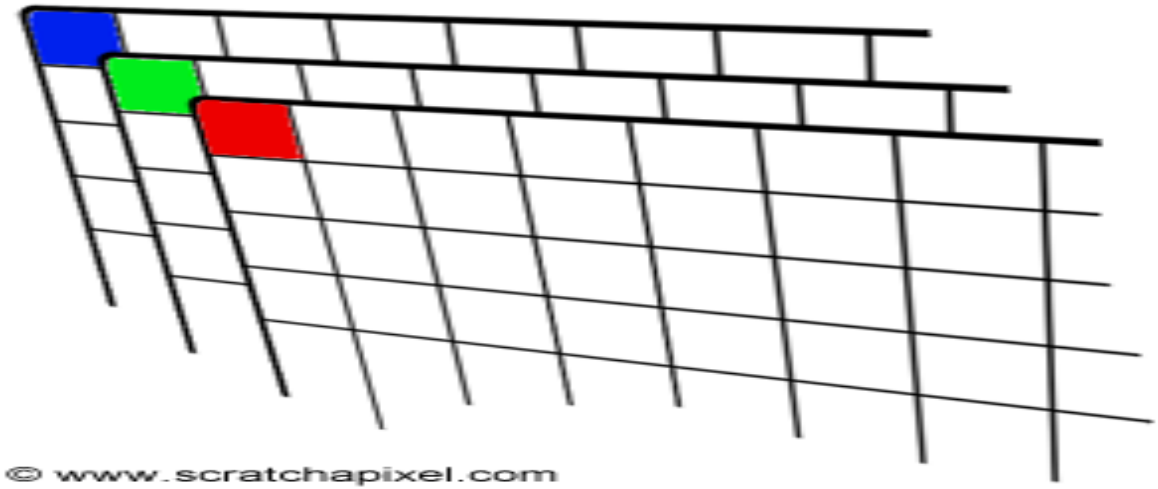
- Portable Network Graphics (PNG): stores images pixel by pixel.
- Structured Query Language (SQL): used to access/modify relational databases.

Color By Numbers...



- Images are a grid of pixels.

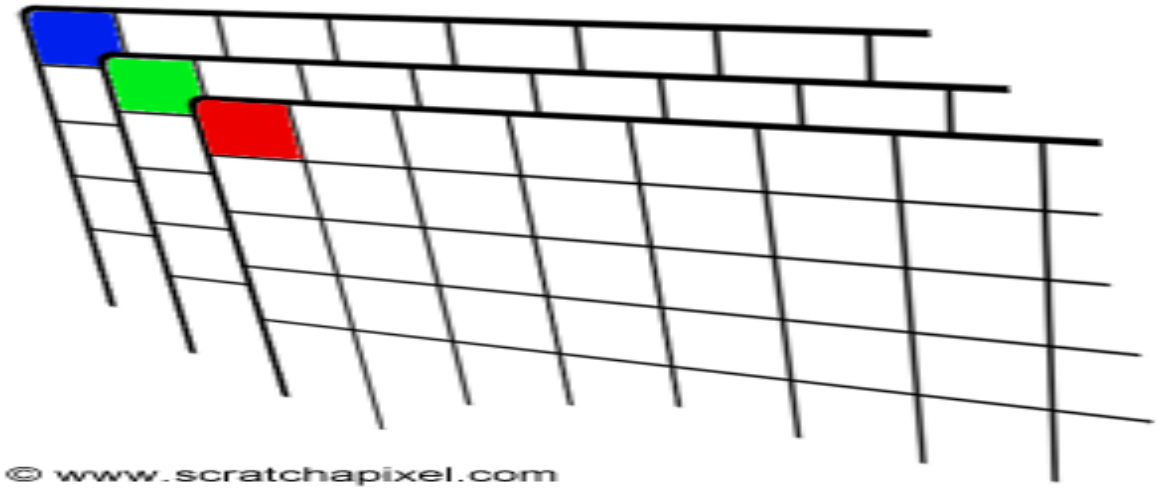
Color By Numbers...



- Images are a grid of pixels.
- Can access color value for pixel (row, col):

```
print "Red:   ", img[row,col,0]
print "Green:", img[row,col,1]
print "Blue:  ", img[row,col,2]
```

Color By Numbers...



- Images are a grid of pixels.
- Can access color value for pixel (row, col):

```
print "Red:   ", img[row,col,0]
print "Green:", img[row,col,1]
print "Blue:  ", img[row,col,2]
```

- Challenges: manipulate images, pixel by pixel.

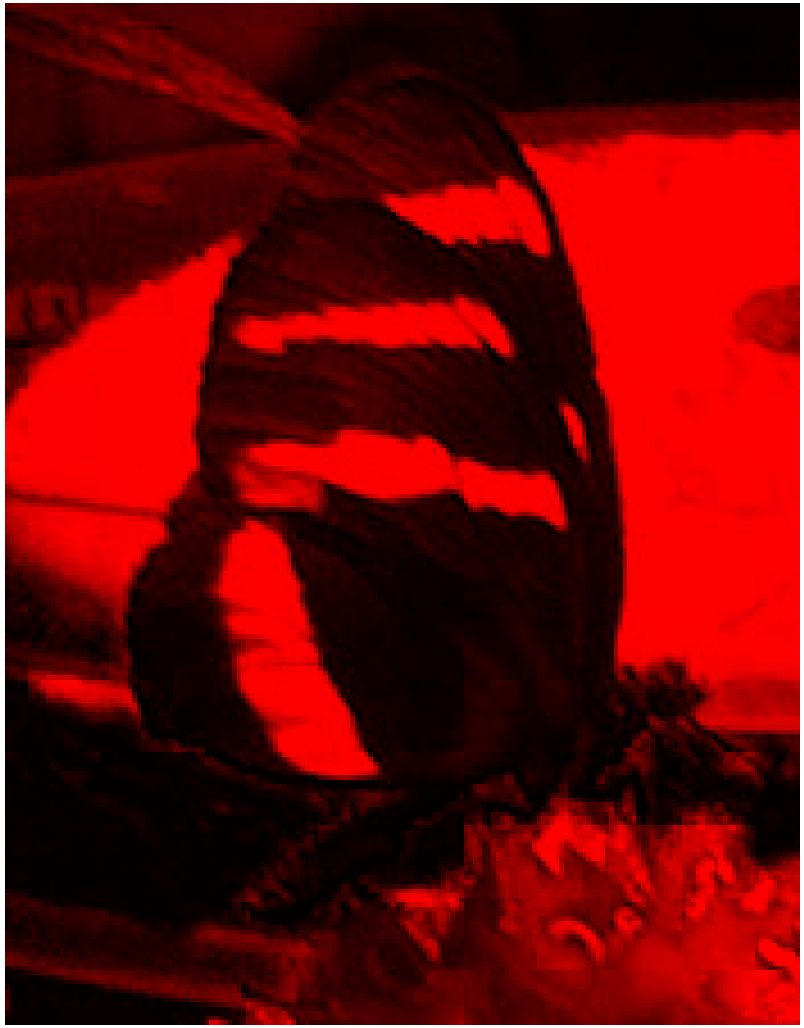
Code Review

filterImage.py - /Users/stjohn/gitHub/bridgeUp-STEM-Miscellanea/imageProcessing/filterImage.py (2.7.9)

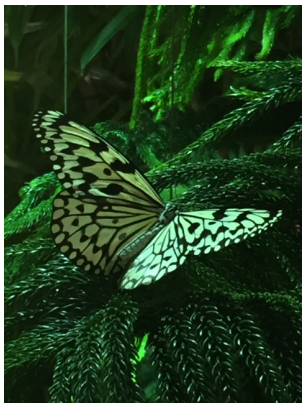
```
#Sample image processing file for BridgeUp: STEM
#Assumes a input file called:  bf1.png is in the same directory

import matplotlib.pyplot as plt    #Import plotting and numbers libraries
import numpy as np

img = plt.imread('bf1.png')    #Read in first butterfly image
plt.imshow(img)                #Load image into matplotlib
plt.show()                     #Open window to show image (close to continue

height = img.shape[0]
width = img.shape[1]
newImg = np.zeros((height,width,3))

for i in range(height):
     for j in range(width):
          newImg[i,j,0] = img[i,j,0]    #Line to change!
          newImg[i,j,1] = 0
          newImg[i,j,2] = 0

plt.imshow(newImg)
plt.show()
plt.imsave('rogue.png',newImg)
```

Ln: 24 Col: 0

Code Review

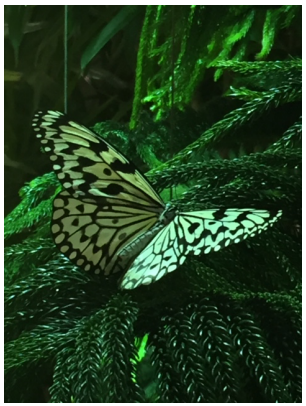


Challenges



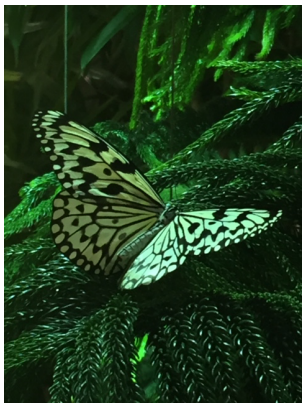
- Display the green channel, the blue channel, and in black and white.

Challenges



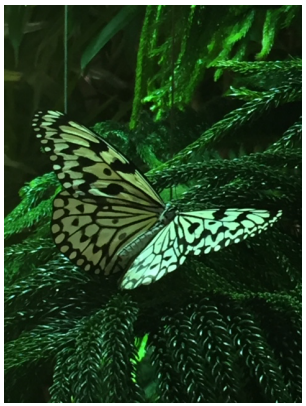
- Display the green channel, the blue channel, and in black and white.
- Starting with a monochromatic butterfly (e.g. `bflyBlue.png`), display only the butterfly.

Challenges



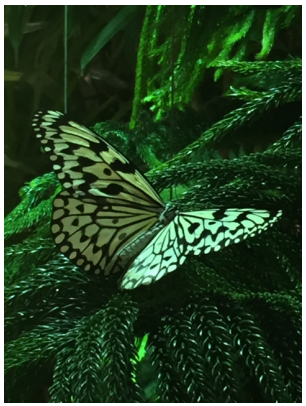
- Display the green channel, the blue channel, and in black and white.
- Starting with a monochromatic butterfly (e.g. `bflyBlue.png`), display only the butterfly.
- Make a white butterfly appear purple (e.g. `bflyGreen.png`).

Challenges



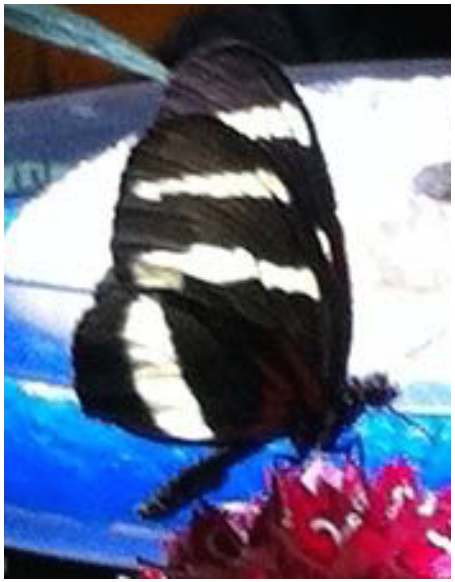
- Display the green channel, the blue channel, and in black and white.
- Starting with a monochromatic butterfly (e.g. `bflyBlue.png`), display only the butterfly.
- Make a white butterfly appear purple (e.g. `bflyGreen.png`).
- Display the image upside down.

Challenges



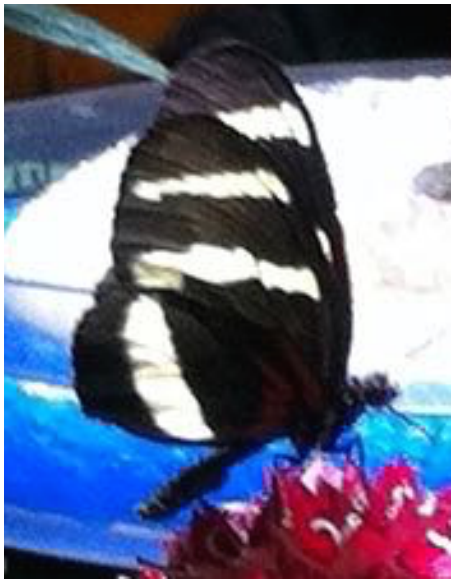
- Display the green channel, the blue channel, and in black and white.
- Starting with a monochromatic butterfly (e.g. `bflyBlue.png`), display only the butterfly.
- Make a white butterfly appear purple (e.g. `bflyGreen.png`).
- Display the image upside down.
- Subtract two similar images and display pixels that differ by more than 30%.

Recap



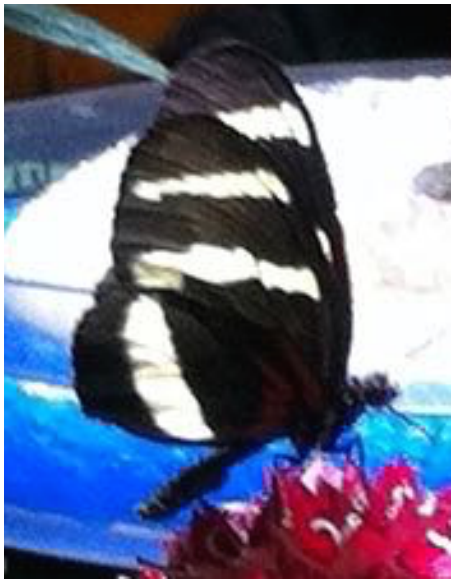
- Using `matplotlib & numpy` in lab today.

Recap



- Using `matplotlib & numpy` in lab today.
- Email lab reports to `kstjohn@amnh.org`

Recap



- Using `matplotlib & numpy` in lab today.
- Email lab reports to `kstjohn@amnh.org`
- Challenges available at `rosalind.info`