

pandas

October 29, 2015

1 Getting Started with Exploratory Data Analysis

3 important Python packages 1. NumPy for efficient computation on arrays 2. Pandas for data analysis 3. Matplotlib for plotting in the notebook

```
In [4]: import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

1.1 Pandas

Python module for manipulating tabular data

1.2 pandas

- Provides python a **DataFrame**
- Structured manipulation tools
- Built on top of **numpy**
- Huge growth from 2011-2012
- Very **efficient**
- Great for *medium* data

Resources

- pandas.pydata.org
- [Python for Data Analysis](#) by Wes McKinney
- [Data Wrangling Kung Fu with Pandas](#) by Wes McKinney
- [Cheat sheet](#) by Quandl

1.2.1 Why pandas?

80% of the effort in data analysis is spent cleaning data. [Hadley Wickham](#)

Efficiency

- Different views of data
- [Tidy data](#) by Hadley Wickham

Raw data is often in the wrong format

- How often to you download an array ready for array-oriented computing?
- e.g. `scikit-learn` interface

Storage may be best in a different format

- Sparse representations
- Upload to database

1.3 Simple example using *seal* observational data

Data from:

- Baker L, Flemming JEM, Jonsen ID, Lidgard DC, Iverson SJ, Bowen WD (2015) A novel approach to quantifying the spatiotemporal behavior of instrumented grey seals used to sample the environment. *Movement Ecology* 3(1):20. doi:10.1186/s40462-015-0047-4
- Lidgard DC, Bowen WD, Iverson SJ (2015) Data from: A novel approach to quantifying the spatiotemporal behavior of instrumented grey seals used to sample the environment. Movebank Data Repository. doi:10.5441/001/1.910p0c20

1.3.1 Reading a CSV file as text

```
In [2]: f = open("Grey seals (Halichoerus grypus) at Sable Island (data from Baker et al. 2015).csv", 'r')
        lines = f.readlines()
        lines[:10]
```

```
Out[2]: ['event-id,visible,timestamp,location-long,location-lat,manually-marked-outlier,sensor-type,ind
'677436629,true,2011-06-15 17:35:18.000,-59.97949982,43.92495728,,,"gps","Halichoerus grypus",
'677436630,true,2011-06-15 17:50:19.000,-59.98273849,43.92548752,,,"gps","Halichoerus grypus",
'677436631,true,2011-06-15 18:05:32.000,-59.98968887,43.92582703,,,"gps","Halichoerus grypus",
'677436632,true,2011-06-15 18:21:27.000,-59.99033737,43.92613602,,,"gps","Halichoerus grypus",
'677436633,true,2011-06-15 18:36:31.000,-59.9889679,43.92525482,,,"gps","Halichoerus grypus",
'677436634,true,2011-06-15 18:51:23.000,-59.98394394,43.92564011,,,"gps","Halichoerus grypus",
'677436635,true,2011-06-15 19:06:20.000,-59.98566055,43.92499924,,,"gps","Halichoerus grypus",
'677436636,true,2011-06-15 19:22:18.000,-59.987854,43.92406082,,,"gps","Halichoerus grypus",
'677436637,true,2011-06-15 19:37:18.000,-59.98072815,43.92603302,,,"gps","Halichoerus grypus",
```

1.3.2 Creating a DataFrame

```
df = pd.read_csv(filename)
print df
```

Why store it this way?

- Similar to a table
- powerfull way to interact with the data

1.3.3 Converting the timestamp Column

NumPy datetime64 dtype

```
In [5]: ?pd.read_csv
```

```
In [36]: df = pd.read_csv("Grey seals (Halichoerus grypus) at Sable Island (data from Baker et al. 2015).csv",
                        parse_dates=[2])
df.head(3)
```

```

Out[36]:      event-id visible      timestamp location-long location-lat \
0  677436629    True 2011-06-15 17:35:18    -59.979500    43.924957
1  677436630    True 2011-06-15 17:50:19    -59.982738    43.925488
2  677436631    True 2011-06-15 18:05:32    -59.989689    43.925827

      manually-marked-outlier sensor-type individual-taxon-canonical-name \
0                                NaN      gps      Halichoerus grypus
1                                NaN      gps      Halichoerus grypus
2                                NaN      gps      Halichoerus grypus

      tag-local-identifier individual-local-identifier \
0                106705                E 87
1                106705                E 87
2                106705                E 87

                                study-name
0  Grey seals (Halichoerus grypus) at Sable Islan...
1  Grey seals (Halichoerus grypus) at Sable Islan...
2  Grey seals (Halichoerus grypus) at Sable Islan...

```

1.3.4 What data types are in the data frame?

```
In [65]: df.dtypes
```

```

Out[65]: event-id      int64
visible      bool
timestamp      datetime64[ns]
location-long      float64
location-lat      float64
manually-marked-outlier      object
sensor-type      object
individual-taxon-canonical-name      object
tag-local-identifier      int64
individual-local-identifier      object
study-name      object
dtype: object

```

1.3.5 Overview over the numerical types of the data

```
In [66]: df.describe()
```

```

Out[66]:      event-id      visible location-long location-lat \
count  1.244350e+05    124435  124435.000000  124435.000000
mean   6.455033e+08    0.999397   -60.352127    44.453460
std    8.829974e+06    0.0245431    1.124249    0.837039
min    6.430037e+08      False   -75.647453    43.408535
25%    6.430355e+08         1    -60.473623    43.928371
50%    6.430666e+08         1    -60.064209    44.198532
75%    6.430977e+08         1    -59.855791    44.689791
max    6.774455e+08         True   -50.937111    59.374504

      tag-local-identifier
count      124435.000000
mean      106714.273621
std         5.576758

```

```

min          106705.000000
25%          106710.000000
50%          106715.000000
75%          106718.000000
max          106724.000000

```

1.3.6 Size of the data set

```
In [7]: len(df)
```

```
Out[7]: 124435
```

1.3.7 Indexing - very similar to Numpy arrays

- 0 based indexing
- last element not included

```
In [8]: df[2:10]
```

```

Out[8]:
   event-id visible      timestamp  location-long  location-lat \
2  677436631    True  2011-06-15 18:05:32    -59.989689    43.925827
3  677436632    True  2011-06-15 18:21:27    -59.990337    43.926136
4  677436633    True  2011-06-15 18:36:31    -59.988968    43.925255
5  677436634    True  2011-06-15 18:51:23    -59.983944    43.925640
6  677436635    True  2011-06-15 19:06:20    -59.985661    43.924999
7  677436636    True  2011-06-15 19:22:18    -59.987854    43.924061
8  677436637    True  2011-06-15 19:37:18    -59.980728    43.926033
9  677436638    True  2011-06-15 22:27:41    -59.981052    43.925884

   manually-marked-outlier  sensor-type  individual-taxon-canonical-name \
2                        NaN          gps      Halichoerus grypus
3                        NaN          gps      Halichoerus grypus
4                        NaN          gps      Halichoerus grypus
5                        NaN          gps      Halichoerus grypus
6                        NaN          gps      Halichoerus grypus
7                        NaN          gps      Halichoerus grypus
8                        NaN          gps      Halichoerus grypus
9                        NaN          gps      Halichoerus grypus

   tag-local-identifier  individual-local-identifier \
2                106705                E 87
3                106705                E 87
4                106705                E 87
5                106705                E 87
6                106705                E 87
7                106705                E 87
8                106705                E 87
9                106705                E 87

   study-name
2  Grey seals (Halichoerus grypus) at Sable Islan...
3  Grey seals (Halichoerus grypus) at Sable Islan...
4  Grey seals (Halichoerus grypus) at Sable Islan...
5  Grey seals (Halichoerus grypus) at Sable Islan...
6  Grey seals (Halichoerus grypus) at Sable Islan...

```

```

7 Grey seals (Halichoerus grypus) at Sable Islan...
8 Grey seals (Halichoerus grypus) at Sable Islan...
9 Grey seals (Halichoerus grypus) at Sable Islan...

```

```
In [6]: df[-5:]
```

```

Out[6]:      event-id visible      timestamp  location-long  location-lat \
124430  643119961     True 2012-01-06 03:16:42    -59.756832    43.974522 \
124431  643119962     True 2012-01-06 03:33:53    -59.757130    43.974274 \
124432  643119963     True 2012-01-06 03:52:42    -59.758202    43.970879 \
124433  643119964     True 2012-01-06 04:10:06    -59.761147    43.971104 \
124434  643119965     True 2012-01-06 04:25:06    -59.762070    43.971161 \

      manually-marked-outlier sensor-type individual-taxon-canonical-name \
124430                      NaN      gps      Halichoerus grypus \
124431                      NaN      gps      Halichoerus grypus \
124432                      NaN      gps      Halichoerus grypus \
124433                      NaN      gps      Halichoerus grypus \
124434                      NaN      gps      Halichoerus grypus \

      tag-local-identifier individual-local-identifier \
124430                106724                F357 \
124431                106724                F357 \
124432                106724                F357 \
124433                106724                F357 \
124434                106724                F357 \

      study-name
124430 Grey seals (Halichoerus grypus) at Sable Islan...
124431 Grey seals (Halichoerus grypus) at Sable Islan...
124432 Grey seals (Halichoerus grypus) at Sable Islan...
124433 Grey seals (Halichoerus grypus) at Sable Islan...
124434 Grey seals (Halichoerus grypus) at Sable Islan...

```

1.3.8 Extracting the values of a column

```

In [15]: longitude = df['location-long'].values
         print(type(longitude))
         longitude

```

```
<type 'numpy.ndarray'>
```

```

Out[15]: array([-59.97949982, -59.98273849, -59.98968887, ..., -59.7582016 ,
               -59.76114655, -59.7620697 ])

```

1.4 Finding how many individuals are tracked

```
In [7]: df["individual-local-identifier"].unique()
```

```

Out[7]: array(['E 87', 'S0749', 'S0757', 'F104', 'S0753', 'F122', 'K 88', 'K 11',
              'S0751', 'S0758', 'S0756', 'F532', 'F719', 'F367', 'F357'], dtype=object)

```

1.5 Extracting columns for a new data frame

```
In [24]: df.head(2)
```

```

Out[24]:      event-id visible      timestamp location-long location-lat \
0  677436629     True 2011-06-15 17:35:18    -59.979500    43.924957
1  677436630     True 2011-06-15 17:50:19    -59.982738    43.925488

      manually-marked-outlier sensor-type individual-taxon-canonical-name \
0                                NaN      gps      Halichoerus grypus
1                                NaN      gps      Halichoerus grypus

      tag-local-identifier individual-local-identifier \
0                106705                        E 87
1                106705                        E 87

                                study-name
0  Grey seals (Halichoerus grypus) at Sable Islan...
1  Grey seals (Halichoerus grypus) at Sable Islan...

```

```

In [22]: sdf = df[["timestamp","location-long","location-lat","individual-local-identifier","event-id"]]
sdf.head(5)

```

```

Out[22]:      timestamp location-long location-lat \
0 2011-06-15 17:35:18    -59.979500    43.924957
1 2011-06-15 17:50:19    -59.982738    43.925488
2 2011-06-15 18:05:32    -59.989689    43.925827
3 2011-06-15 18:21:27    -59.990337    43.926136
4 2011-06-15 18:36:31    -59.988968    43.925255

      individual-local-identifier event-id
0                E 87  677436629
1                E 87  677436630
2                E 87  677436631
3                E 87  677436632
4                E 87  677436633

```

1.6 Using the timestamp as an index

```

In [25]: sdf.set_index("timestamp",inplace=True)
sdf.head(5)

```

```

Out[25]:      location-long location-lat individual-local-identifier \
timestamp
2011-06-15 17:35:18    -59.979500    43.924957                E 87
2011-06-15 17:50:19    -59.982738    43.925488                E 87
2011-06-15 18:05:32    -59.989689    43.925827                E 87
2011-06-15 18:21:27    -59.990337    43.926136                E 87
2011-06-15 18:36:31    -59.988968    43.925255                E 87

      event-id
timestamp
2011-06-15 17:35:18  677436629
2011-06-15 17:50:19  677436630
2011-06-15 18:05:32  677436631
2011-06-15 18:21:27  677436632
2011-06-15 18:36:31  677436633

```

1.7 Adding a column

- Want to add a behavior index to the data.
- Data is in a numpy array

```
In [26]: behav = np.random.randn(len(sdf))
         sdf.insert(4, 'behavior', behav)
         sdf.head(5)
```

```
Out[26]:
```

		location-long	location-lat	individual-local-identifier	\
	timestamp				
	2011-06-15 17:35:18	-59.979500	43.924957		E 87
	2011-06-15 17:50:19	-59.982738	43.925488		E 87
	2011-06-15 18:05:32	-59.989689	43.925827		E 87
	2011-06-15 18:21:27	-59.990337	43.926136		E 87
	2011-06-15 18:36:31	-59.988968	43.925255		E 87

		event-id	behavior
	timestamp		
	2011-06-15 17:35:18	677436629	-0.283397
	2011-06-15 17:50:19	677436630	-1.784967
	2011-06-15 18:05:32	677436631	-1.428955
	2011-06-15 18:21:27	677436632	0.160293
	2011-06-15 18:36:31	677436633	1.749530

1.8 Renaming columns

- location-long -> longitude
- location-lat -> latitude
- individual-local-identifier -> individual

```
In [27]: sdf = sdf.rename(columns={"location-long": "longitude",
                                   "location-lat": "latitude",
                                   "individual-local-identifier": "individual"})
         sdf.head(5)
```

```
Out[27]:
```

		longitude	latitude	individual	event-id	behavior
	timestamp					
	2011-06-15 17:35:18	-59.979500	43.924957	E 87	677436629	-0.283397
	2011-06-15 17:50:19	-59.982738	43.925488	E 87	677436630	-1.784967
	2011-06-15 18:05:32	-59.989689	43.925827	E 87	677436631	-1.428955
	2011-06-15 18:21:27	-59.990337	43.926136	E 87	677436632	0.160293
	2011-06-15 18:36:31	-59.988968	43.925255	E 87	677436633	1.749530

1.9 Writing to a csv file

```
In [28]: sdf.to_csv("seal-behav.csv")
         !head "seal-behav.csv"
```

```
timestamp,longitude,latitude,individual,event-id,behavior
2011-06-15 17:35:18,-59.97949982,43.92495728,E 87,677436629,-0.283397437027
2011-06-15 17:50:19,-59.98273849,43.92548752,E 87,677436630,-1.78496655764
2011-06-15 18:05:32,-59.98968887,43.92582703,E 87,677436631,-1.42895470333
2011-06-15 18:21:27,-59.99033737,43.92613602,E 87,677436632,0.160293467809
2011-06-15 18:36:31,-59.9889679,43.92525482,E 87,677436633,1.74952967386
2011-06-15 18:51:23,-59.98394394,43.92564011,E 87,677436634,-0.256575728289
```

```

2011-06-15 19:06:20,-59.98566055,43.92499924,E 87,677436635,1.07774828185
2011-06-15 19:22:18,-59.987854,43.92406082,E 87,677436636,-1.91521313598
2011-06-15 19:37:18,-59.98072815,43.92603302,E 87,677436637,-0.499820308692

```

2 Hierarchical columns

2.1 Reorder the organization of the table

- Index: timestamp
- Columns: the individuals

```

In [29]: sd = sdf.pivot(columns='individual') #row, column, values (optional)
sd[:5]

```

```

Out[29]:
          longitude
individual      E 87 F104 F122 F357 F367 F532 F719 K 11 K 88
timestamp
2011-06-11 19:07:27      NaN NaN NaN NaN NaN NaN NaN NaN NaN
2011-06-11 19:24:28      NaN NaN NaN NaN NaN NaN NaN NaN NaN
2011-06-11 19:41:25      NaN NaN NaN NaN NaN NaN NaN NaN NaN
2011-06-11 19:57:39      NaN NaN NaN NaN NaN NaN NaN NaN NaN
2011-06-11 20:14:14      NaN NaN NaN NaN NaN NaN NaN NaN NaN

```

```

          ... behavior
individual      S0749 ...      F532 F719 K 11 K 88      S0749 S0751
timestamp      ...
2011-06-11 19:07:27 -59.961720 ...      NaN NaN NaN NaN 2.066740 NaN
2011-06-11 19:24:28 -59.960075 ...      NaN NaN NaN NaN 0.805529 NaN
2011-06-11 19:41:25 -59.956333 ...      NaN NaN NaN NaN 0.232823 NaN
2011-06-11 19:57:39 -59.957340 ...      NaN NaN NaN NaN 0.261920 NaN
2011-06-11 20:14:14 -59.965260 ...      NaN NaN NaN NaN -0.098277 NaN

```

```

individual      S0753 S0756 S0757 S0758
timestamp
2011-06-11 19:07:27      NaN NaN NaN NaN
2011-06-11 19:24:28      NaN NaN NaN NaN
2011-06-11 19:41:25      NaN NaN NaN NaN
2011-06-11 19:57:39      NaN NaN NaN NaN
2011-06-11 20:14:14      NaN NaN NaN NaN

```

[5 rows x 60 columns]

```

In [61]: sd['behavior'][:5]

```

```

Out[61]:
individual      E 87 F104 F122 F357 F367 F532 F719 K 11 K 88
timestamp
2011-06-11 19:07:27      NaN NaN NaN NaN NaN NaN NaN NaN NaN
2011-06-11 19:24:28      NaN NaN NaN NaN NaN NaN NaN NaN NaN
2011-06-11 19:41:25      NaN NaN NaN NaN NaN NaN NaN NaN NaN
2011-06-11 19:57:39      NaN NaN NaN NaN NaN NaN NaN NaN NaN
2011-06-11 20:14:14      NaN NaN NaN NaN NaN NaN NaN NaN NaN

```

```

individual      S0749 S0751 S0753 S0756 S0757 S0758
timestamp

```


2011-06-11	19:07:27	0.408271	NaN	NaN	NaN	NaN	NaN
2011-06-11	19:24:28	1.446369	NaN	NaN	NaN	NaN	NaN
2011-06-11	19:41:25	-1.080983	NaN	NaN	NaN	NaN	NaN
2011-06-11	19:57:39	0.060002	NaN	NaN	NaN	NaN	NaN
2011-06-11	20:14:14	0.685842	NaN	NaN	NaN	NaN	NaN

2.2 Extracting longitude and latitude for each individual

```
In [30]: longLat = sdf[['individual', 'longitude', 'latitude']]
         longLat[2::5000]
```

```
Out[30]:
```

	timestamp	individual	longitude	latitude
	2011-06-15 18:05:32	E 87	-59.989689	43.925827
	2011-09-26 15:03:18	E 87	-60.023972	43.928467
	2011-07-01 14:24:30	S0749	-59.932743	44.020000
	2011-10-17 15:09:15	S0749	-60.100521	44.495800
	2011-07-07 01:14:48	S0757	-59.275227	44.682209
	2011-12-01 04:11:42	S0757	-60.867977	44.248436
	2011-08-20 09:16:11	F104	-60.041412	43.927155
	2011-11-16 19:40:27	F104	-60.496716	44.621536
	2011-07-20 04:25:05	S0753	-59.538494	44.334286
	2011-10-12 08:00:38	S0753	-59.769253	44.785385
	2011-06-23 10:07:36	F122	-59.639961	44.038654
	2011-08-09 14:19:54	K 88	-59.081959	43.875942
	2011-11-18 02:43:41	K 88	-61.186897	43.768154
	2011-08-21 08:05:02	K 11	-60.871934	45.303607
	2011-12-05 10:54:25	K 11	-60.863003	44.177757
	2011-09-05 20:35:02	S0751	-64.866554	46.786152
	2011-07-15 15:28:05	S0758	-60.021687	43.928288
	2011-10-27 03:14:26	S0758	-59.938644	43.758900
	2011-07-05 18:11:54	S0756	-59.991898	44.616425
	2011-10-16 07:19:25	S0756	-60.324448	45.068775
	2011-06-17 14:07:51	F532	-60.800690	43.785324
	2011-12-24 18:04:44	F532	-60.248558	43.836136
	2011-07-02 12:27:12	F367	-60.241093	44.314720
	2011-08-13 11:41:55	F357	-60.123775	43.957939
	2011-10-28 19:59:47	F357	-60.425831	45.067749

2.2.1 Same for the pivoted data

```
In [32]: sd[['longitude', 'latitude']][:500]
```

```
Out[32]:
```

	individual	longitude				
		E 87	F104	F122	F357	F367
	timestamp					
	2011-06-11 19:07:27	NaN	NaN	NaN	NaN	NaN
	2011-06-14 15:26:02	NaN	NaN	NaN	NaN	NaN
	2011-06-15 15:37:00	NaN	NaN	NaN	NaN	NaN
	2011-06-16 19:26:59	NaN	NaN	NaN	NaN	NaN
	2011-06-17 17:36:40	NaN	NaN	NaN	NaN	NaN
	2011-06-18 09:37:14	NaN	NaN	NaN	NaN	NaN
	2011-06-19 00:50:08	NaN	NaN	NaN	NaN	NaN
	2011-06-19 15:40:35	NaN	NaN	NaN	-60.376617	NaN
	2011-06-20 09:54:23	NaN	NaN	NaN	NaN	NaN

2011-06-21 06:58:05	NaN	NaN	NaN	NaN	NaN
2011-06-22 03:16:32	NaN	NaN	NaN	NaN	NaN
2011-06-22 19:11:02	NaN	-59.923889	NaN	NaN	NaN
2011-06-23 11:21:56	NaN	NaN	NaN	NaN	NaN
2011-06-24 05:19:18	-59.818169	NaN	NaN	NaN	NaN
2011-06-25 00:17:55	NaN	NaN	NaN	NaN	NaN
2011-06-25 17:06:53	NaN	NaN	NaN	NaN	NaN
2011-06-26 08:03:44	NaN	NaN	NaN	NaN	NaN
2011-06-27 00:56:59	NaN	NaN	NaN	NaN	NaN
2011-06-27 14:40:18	NaN	NaN	NaN	NaN	NaN
2011-06-28 08:20:12	NaN	NaN	NaN	-60.160671	NaN
2011-06-29 01:15:33	NaN	NaN	NaN	NaN	NaN
2011-06-29 21:41:18	NaN	NaN	-59.557892	NaN	NaN
2011-06-30 15:40:20	NaN	NaN	NaN	NaN	NaN
2011-07-01 09:24:32	NaN	NaN	NaN	NaN	NaN
2011-07-02 03:32:47	NaN	NaN	NaN	NaN	NaN
2011-07-02 20:03:28	NaN	NaN	NaN	NaN	NaN
2011-07-03 11:16:26	NaN	NaN	NaN	NaN	NaN
2011-07-04 05:50:32	NaN	-60.011074	NaN	NaN	NaN
2011-07-04 23:14:22	NaN	NaN	NaN	-60.214314	NaN
2011-07-05 17:40:17	NaN	NaN	NaN	NaN	NaN
...
2011-12-04 10:41:44	NaN	NaN	NaN	NaN	NaN
2011-12-05 03:15:54	NaN	NaN	NaN	NaN	NaN
2011-12-05 20:01:00	NaN	NaN	NaN	NaN	NaN
2011-12-06 13:23:54	NaN	NaN	NaN	NaN	NaN
2011-12-07 07:13:25	NaN	NaN	NaN	NaN	NaN
2011-12-07 23:21:48	NaN	NaN	NaN	NaN	NaN
2011-12-08 19:31:39	NaN	NaN	NaN	NaN	NaN
2011-12-09 15:52:20	NaN	NaN	NaN	NaN	NaN
2011-12-10 12:25:36	NaN	NaN	NaN	NaN	NaN
2011-12-11 09:33:03	NaN	NaN	NaN	NaN	NaN
2011-12-12 04:38:13	NaN	NaN	NaN	NaN	NaN
2011-12-13 01:49:53	NaN	NaN	NaN	NaN	NaN
2011-12-13 20:30:13	-59.780468	NaN	NaN	NaN	NaN
2011-12-14 13:05:34	NaN	NaN	NaN	NaN	NaN
2011-12-15 05:50:24	NaN	NaN	NaN	NaN	NaN
2011-12-15 23:04:30	NaN	NaN	NaN	NaN	NaN
2011-12-16 16:46:05	NaN	NaN	NaN	NaN	NaN
2011-12-17 10:42:00	NaN	NaN	NaN	NaN	NaN
2011-12-18 03:30:04	NaN	NaN	NaN	-60.422989	NaN
2011-12-18 21:26:41	NaN	NaN	NaN	-60.424610	NaN
2011-12-19 19:34:50	NaN	NaN	NaN	NaN	NaN
2011-12-20 18:14:31	NaN	NaN	NaN	NaN	NaN
2011-12-21 11:01:48	-59.963799	NaN	NaN	NaN	NaN
2011-12-22 08:12:18	-59.735023	NaN	NaN	NaN	NaN
2011-12-23 07:38:43	NaN	NaN	NaN	NaN	NaN
2011-12-24 05:06:45	NaN	NaN	NaN	NaN	NaN
2011-12-25 12:12:54	NaN	NaN	NaN	-60.423866	NaN
2011-12-27 09:44:07	NaN	NaN	NaN	NaN	NaN
2011-12-31 05:34:22	NaN	NaN	NaN	NaN	NaN
2012-01-17 02:29:32	-50.937111	NaN	NaN	NaN	NaN

\

individual		F532	F719	K 11	K 88	S0749
timestamp						
2011-06-11 19:07:27		NaN	NaN	NaN	NaN	-59.961720
2011-06-14 15:26:02		NaN	NaN	NaN	NaN	NaN
2011-06-15 15:37:00		NaN	NaN	NaN	NaN	NaN
2011-06-16 19:26:59		NaN	NaN	NaN	NaN	-60.129692
2011-06-17 17:36:40		NaN	NaN	NaN	NaN	NaN
2011-06-18 09:37:14	-60.743782	NaN	NaN	NaN	NaN	NaN
2011-06-19 00:50:08		NaN	NaN	NaN	NaN	NaN
2011-06-19 15:40:35		NaN	NaN	NaN	NaN	NaN
2011-06-20 09:54:23		NaN	NaN	NaN	NaN	NaN
2011-06-21 06:58:05		NaN	NaN	NaN	NaN	NaN
2011-06-22 03:16:32		NaN	NaN	NaN	NaN	NaN
2011-06-22 19:11:02		NaN	NaN	NaN	NaN	NaN
2011-06-23 11:21:56		NaN	NaN	NaN	NaN	NaN
2011-06-24 05:19:18		NaN	NaN	NaN	NaN	NaN
2011-06-25 00:17:55		NaN	NaN	NaN	NaN	NaN
2011-06-25 17:06:53		NaN	NaN	NaN	NaN	NaN
2011-06-26 08:03:44		NaN	NaN	NaN	NaN	-60.083836
2011-06-27 00:56:59		NaN	NaN	NaN	NaN	NaN
2011-06-27 14:40:18		NaN	NaN	-59.890358	NaN	NaN
2011-06-28 08:20:12		NaN	NaN	NaN	NaN	NaN
2011-06-29 01:15:33		NaN	NaN	NaN	NaN	NaN
2011-06-29 21:41:18		NaN	NaN	NaN	NaN	NaN
2011-06-30 15:40:20		NaN	NaN	NaN	NaN	NaN
2011-07-01 09:24:32		NaN	NaN	NaN	NaN	NaN
2011-07-02 03:32:47		NaN	NaN	-59.217457	NaN	NaN
2011-07-02 20:03:28		NaN	NaN	NaN	NaN	NaN
2011-07-03 11:16:26		NaN	NaN	NaN	NaN	-60.036774
2011-07-04 05:50:32		NaN	NaN	NaN	NaN	NaN
2011-07-04 23:14:22		NaN	NaN	NaN	NaN	NaN
2011-07-05 17:40:17		NaN	NaN	NaN	NaN	NaN
...	
2011-12-04 10:41:44		NaN	NaN	NaN	NaN	NaN
2011-12-05 03:15:54		NaN	NaN	-60.880478	NaN	NaN
2011-12-05 20:01:00	-61.250111	NaN	NaN	NaN	NaN	NaN
2011-12-06 13:23:54		NaN	NaN	-60.875164	NaN	NaN
2011-12-07 07:13:25		NaN	NaN	NaN	NaN	NaN
2011-12-07 23:21:48		NaN	NaN	NaN	NaN	-59.966709
2011-12-08 19:31:39		NaN	NaN	NaN	NaN	NaN
2011-12-09 15:52:20		NaN	NaN	NaN	NaN	NaN
2011-12-10 12:25:36		NaN	NaN	NaN	NaN	NaN
2011-12-11 09:33:03		NaN	NaN	NaN	NaN	NaN
2011-12-12 04:38:13		NaN	NaN	NaN	-61.275227	NaN
2011-12-13 01:49:53		NaN	NaN	NaN	NaN	NaN
2011-12-13 20:30:13		NaN	NaN	NaN	NaN	NaN
2011-12-14 13:05:34		NaN	NaN	NaN	-61.320328	NaN
2011-12-15 05:50:24		NaN	NaN	NaN	-61.245434	NaN
2011-12-15 23:04:30		NaN	NaN	NaN	NaN	NaN
2011-12-16 16:46:05		NaN	NaN	NaN	-61.311745	NaN
2011-12-17 10:42:00	-60.672268	NaN	NaN	NaN	NaN	NaN
2011-12-18 03:30:04		NaN	NaN	NaN	NaN	NaN
2011-12-18 21:26:41		NaN	NaN	NaN	NaN	NaN
2011-12-19 19:34:50		NaN	NaN	NaN	NaN	NaN

2011-12-20	18:14:31	NaN	NaN	-60.806515	NaN	NaN
2011-12-21	11:01:48	NaN	NaN	NaN	NaN	NaN
2011-12-22	08:12:18	NaN	NaN	NaN	NaN	NaN
2011-12-23	07:38:43	NaN	NaN	NaN	NaN	NaN
2011-12-24	05:06:45	NaN	NaN	NaN	NaN	NaN
2011-12-25	12:12:54	NaN	NaN	NaN	NaN	NaN
2011-12-27	09:44:07	-59.812279	NaN	NaN	NaN	NaN
2011-12-31	05:34:22	NaN	NaN	NaN	NaN	NaN
2012-01-17	02:29:32	NaN	NaN	NaN	NaN	NaN

	...	latitude			
individual	...	F532 F719	K 11	K 88	
timestamp	...				
2011-06-11	19:07:27	NaN	NaN	NaN	NaN
2011-06-14	15:26:02	NaN	NaN	NaN	NaN
2011-06-15	15:37:00	NaN	NaN	NaN	NaN
2011-06-16	19:26:59	NaN	NaN	NaN	NaN
2011-06-17	17:36:40	NaN	NaN	NaN	NaN
2011-06-18	09:37:14	43.812111	NaN	NaN	NaN
2011-06-19	00:50:08	NaN	NaN	NaN	NaN
2011-06-19	15:40:35	NaN	NaN	NaN	NaN
2011-06-20	09:54:23	NaN	NaN	NaN	NaN
2011-06-21	06:58:05	NaN	NaN	NaN	NaN
2011-06-22	03:16:32	NaN	NaN	NaN	NaN
2011-06-22	19:11:02	NaN	NaN	NaN	NaN
2011-06-23	11:21:56	NaN	NaN	NaN	NaN
2011-06-24	05:19:18	NaN	NaN	NaN	NaN
2011-06-25	00:17:55	NaN	NaN	NaN	NaN
2011-06-25	17:06:53	NaN	NaN	NaN	NaN
2011-06-26	08:03:44	NaN	NaN	NaN	NaN
2011-06-27	00:56:59	NaN	NaN	NaN	NaN
2011-06-27	14:40:18	NaN	NaN	44.111855	NaN
2011-06-28	08:20:12	NaN	NaN	NaN	NaN
2011-06-29	01:15:33	NaN	NaN	NaN	NaN
2011-06-29	21:41:18	NaN	NaN	NaN	NaN
2011-06-30	15:40:20	NaN	NaN	NaN	NaN
2011-07-01	09:24:32	NaN	NaN	NaN	NaN
2011-07-02	03:32:47	NaN	NaN	44.049450	NaN
2011-07-02	20:03:28	NaN	NaN	NaN	NaN
2011-07-03	11:16:26	NaN	NaN	NaN	NaN
2011-07-04	05:50:32	NaN	NaN	NaN	NaN
2011-07-04	23:14:22	NaN	NaN	NaN	NaN
2011-07-05	17:40:17	NaN	NaN	NaN	NaN
...
2011-12-04	10:41:44	NaN	NaN	NaN	NaN
2011-12-05	03:15:54	NaN	NaN	44.168858	NaN
2011-12-05	20:01:00	44.634060	NaN	NaN	NaN
2011-12-06	13:23:54	NaN	NaN	44.173916	NaN
2011-12-07	07:13:25	NaN	NaN	NaN	NaN
2011-12-07	23:21:48	NaN	NaN	NaN	NaN
2011-12-08	19:31:39	NaN	NaN	NaN	NaN
2011-12-09	15:52:20	NaN	NaN	NaN	NaN
2011-12-10	12:25:36	NaN	NaN	NaN	NaN
2011-12-11	09:33:03	NaN	NaN	NaN	NaN

2011-12-12 04:38:13	...	NaN	NaN	NaN	43.782669
2011-12-13 01:49:53	...	NaN	NaN	NaN	NaN
2011-12-13 20:30:13	...	NaN	NaN	NaN	NaN
2011-12-14 13:05:34	...	NaN	NaN	NaN	43.851135
2011-12-15 05:50:24	...	NaN	NaN	NaN	43.813744
2011-12-15 23:04:30	...	NaN	NaN	NaN	NaN
2011-12-16 16:46:05	...	NaN	NaN	NaN	43.861801
2011-12-17 10:42:00	...	43.464920	NaN	NaN	NaN
2011-12-18 03:30:04	...	NaN	NaN	NaN	NaN
2011-12-18 21:26:41	...	NaN	NaN	NaN	NaN
2011-12-19 19:34:50	...	NaN	NaN	NaN	NaN
2011-12-20 18:14:31	...	NaN	NaN	44.659252	NaN
2011-12-21 11:01:48	...	NaN	NaN	NaN	NaN
2011-12-22 08:12:18	...	NaN	NaN	NaN	NaN
2011-12-23 07:38:43	...	NaN	NaN	NaN	NaN
2011-12-24 05:06:45	...	NaN	NaN	NaN	NaN
2011-12-25 12:12:54	...	NaN	NaN	NaN	NaN
2011-12-27 09:44:07	...	43.661274	NaN	NaN	NaN
2011-12-31 05:34:22	...	NaN	NaN	NaN	NaN
2012-01-17 02:29:32	...	NaN	NaN	NaN	NaN

individual timestamp	S0749	S0751	S0753	S0756	S0757
2011-06-11 19:07:27	43.936630	NaN	NaN	NaN	NaN
2011-06-14 15:26:02	NaN	NaN	NaN	NaN	43.923386
2011-06-15 15:37:00	NaN	43.934605	NaN	NaN	NaN
2011-06-16 19:26:59	44.530636	NaN	NaN	NaN	NaN
2011-06-17 17:36:40	NaN	NaN	NaN	44.153141	NaN
2011-06-18 09:37:14	NaN	NaN	NaN	NaN	NaN
2011-06-19 00:50:08	NaN	NaN	NaN	NaN	44.150009
2011-06-19 15:40:35	NaN	NaN	NaN	NaN	NaN
2011-06-20 09:54:23	NaN	NaN	NaN	44.546753	NaN
2011-06-21 06:58:05	NaN	NaN	NaN	NaN	44.148384
2011-06-22 03:16:32	NaN	NaN	NaN	NaN	NaN
2011-06-22 19:11:02	NaN	NaN	NaN	NaN	NaN
2011-06-23 11:21:56	NaN	NaN	NaN	NaN	NaN
2011-06-24 05:19:18	NaN	NaN	NaN	NaN	NaN
2011-06-25 00:17:55	NaN	NaN	NaN	44.434975	NaN
2011-06-25 17:06:53	NaN	NaN	NaN	NaN	NaN
2011-06-26 08:03:44	44.539062	NaN	NaN	NaN	NaN
2011-06-27 00:56:59	NaN	NaN	44.505322	NaN	NaN
2011-06-27 14:40:18	NaN	NaN	NaN	NaN	NaN
2011-06-28 08:20:12	NaN	NaN	NaN	NaN	NaN
2011-06-29 01:15:33	NaN	NaN	NaN	44.672211	NaN
2011-06-29 21:41:18	NaN	NaN	NaN	NaN	NaN
2011-06-30 15:40:20	NaN	NaN	NaN	NaN	NaN
2011-07-01 09:24:32	NaN	47.567276	NaN	NaN	NaN
2011-07-02 03:32:47	NaN	NaN	NaN	NaN	NaN
2011-07-02 20:03:28	NaN	NaN	44.572083	NaN	NaN
2011-07-03 11:16:26	44.470207	NaN	NaN	NaN	NaN
2011-07-04 05:50:32	NaN	NaN	NaN	NaN	NaN
2011-07-04 23:14:22	NaN	NaN	NaN	NaN	NaN
2011-07-05 17:40:17	NaN	NaN	44.574051	NaN	NaN

...
2011-12-04 10:41:44	NaN	NaN	NaN	NaN	NaN
2011-12-05 03:15:54	NaN	NaN	NaN	NaN	NaN
2011-12-05 20:01:00	NaN	NaN	NaN	NaN	NaN
2011-12-06 13:23:54	NaN	NaN	NaN	NaN	NaN
2011-12-07 07:13:25	NaN	NaN	NaN	NaN	44.169579
2011-12-07 23:21:48	44.467358	NaN	NaN	NaN	NaN
2011-12-08 19:31:39	NaN	NaN	NaN	NaN	NaN
2011-12-09 15:52:20	NaN	NaN	NaN	NaN	NaN
2011-12-10 12:25:36	NaN	NaN	NaN	44.674633	NaN
2011-12-11 09:33:03	NaN	NaN	NaN	44.642273	NaN
2011-12-12 04:38:13	NaN	NaN	NaN	NaN	NaN
2011-12-13 01:49:53	NaN	NaN	NaN	NaN	44.117413
2011-12-13 20:30:13	NaN	NaN	NaN	NaN	NaN
2011-12-14 13:05:34	NaN	NaN	NaN	NaN	NaN
2011-12-15 05:50:24	NaN	NaN	NaN	NaN	NaN
2011-12-15 23:04:30	NaN	NaN	NaN	NaN	44.208092
2011-12-16 16:46:05	NaN	NaN	NaN	NaN	NaN
2011-12-17 10:42:00	NaN	NaN	NaN	NaN	NaN
2011-12-18 03:30:04	NaN	NaN	NaN	NaN	NaN
2011-12-18 21:26:41	NaN	NaN	NaN	NaN	NaN
2011-12-19 19:34:50	NaN	NaN	NaN	NaN	44.033161
2011-12-20 18:14:31	NaN	NaN	NaN	NaN	NaN
2011-12-21 11:01:48	NaN	NaN	NaN	NaN	NaN
2011-12-22 08:12:18	NaN	NaN	NaN	NaN	NaN
2011-12-23 07:38:43	NaN	NaN	NaN	NaN	NaN
2011-12-24 05:06:45	NaN	NaN	NaN	NaN	43.937801
2011-12-25 12:12:54	NaN	NaN	NaN	NaN	NaN
2011-12-27 09:44:07	NaN	NaN	NaN	NaN	NaN
2011-12-31 05:34:22	NaN	NaN	NaN	44.133907	NaN
2012-01-17 02:29:32	NaN	NaN	NaN	NaN	NaN

individual	S0758
timestamp	
2011-06-11 19:07:27	NaN
2011-06-14 15:26:02	NaN
2011-06-15 15:37:00	NaN
2011-06-16 19:26:59	NaN
2011-06-17 17:36:40	NaN
2011-06-18 09:37:14	NaN
2011-06-19 00:50:08	NaN
2011-06-19 15:40:35	NaN
2011-06-20 09:54:23	NaN
2011-06-21 06:58:05	NaN
2011-06-22 03:16:32	43.643261
2011-06-22 19:11:02	NaN
2011-06-23 11:21:56	43.620010
2011-06-24 05:19:18	NaN
2011-06-25 00:17:55	NaN
2011-06-25 17:06:53	43.651505
2011-06-26 08:03:44	NaN
2011-06-27 00:56:59	NaN
2011-06-27 14:40:18	NaN

```

2011-06-28 08:20:12      NaN
2011-06-29 01:15:33      NaN
2011-06-29 21:41:18      NaN
2011-06-30 15:40:20  43.926842
2011-07-01 09:24:32      NaN
2011-07-02 03:32:47      NaN
2011-07-02 20:03:28      NaN
2011-07-03 11:16:26      NaN
2011-07-04 05:50:32      NaN
2011-07-04 23:14:22      NaN
2011-07-05 17:40:17      NaN
...
2011-12-04 10:41:44  43.772217
2011-12-05 03:15:54      NaN
2011-12-05 20:01:00      NaN
2011-12-06 13:23:54      NaN
2011-12-07 07:13:25      NaN
2011-12-07 23:21:48      NaN
2011-12-08 19:31:39  43.702202
2011-12-09 15:52:20  43.931297
2011-12-10 12:25:36      NaN
2011-12-11 09:33:03      NaN
2011-12-12 04:38:13      NaN
2011-12-13 01:49:53      NaN
2011-12-13 20:30:13      NaN
2011-12-14 13:05:34      NaN
2011-12-15 05:50:24      NaN
2011-12-15 23:04:30      NaN
2011-12-16 16:46:05      NaN
2011-12-17 10:42:00      NaN
2011-12-18 03:30:04      NaN
2011-12-18 21:26:41      NaN
2011-12-19 19:34:50      NaN
2011-12-20 18:14:31      NaN
2011-12-21 11:01:48      NaN
2011-12-22 08:12:18      NaN
2011-12-23 07:38:43  43.471710
2011-12-24 05:06:45      NaN
2011-12-25 12:12:54      NaN
2011-12-27 09:44:07      NaN
2011-12-31 05:34:22      NaN
2012-01-17 02:29:32      NaN

```

[249 rows x 30 columns]

2.2.2 Extracting Data from a Data Frame with a Condition

- Extracting Seal F719 from the table

```
In [33]: df[df["individual-local-identifier"] == "F719"][:5]
```

```

Out[33]:
   event-id  visible  timestamp  location-long  location-lat \
105100  643100631   True  2011-06-14 19:59:29    -59.826447   43.937832
105101  643100632   True  2011-06-14 20:32:38    -59.829632   43.931831
105102  643100633   True  2011-06-14 20:48:07    -59.813534   43.936577

```

105103	643100634	True	2011-06-14 21:18:03	-59.812885	43.942875
105104	643100635	True	2011-06-14 21:33:03	-59.808041	43.941067

	manually-marked-outlier	sensor-type	individual	taxon	canonical-name
105100	NaN	gps		Halichoerus	grypus
105101	NaN	gps		Halichoerus	grypus
105102	NaN	gps		Halichoerus	grypus
105103	NaN	gps		Halichoerus	grypus
105104	NaN	gps		Halichoerus	grypus

	tag-local-identifier	individual-local-identifier
105100	106721	F719
105101	106721	F719
105102	106721	F719
105103	106721	F719
105104	106721	F719

	study-name
105100	Grey seals (Halichoerus grypus) at Sable Islan...
105101	Grey seals (Halichoerus grypus) at Sable Islan...
105102	Grey seals (Halichoerus grypus) at Sable Islan...
105103	Grey seals (Halichoerus grypus) at Sable Islan...
105104	Grey seals (Halichoerus grypus) at Sable Islan...

```
In [34]: sd[['behavior', "F719"], ('latitude', "F719"), ('longitude', "F719")][[:5]]
```

```
Out[34]:
```

	behavior	latitude	longitude
individual	F719	F719	F719
timestamp			
2011-06-11 19:07:27	NaN	NaN	NaN
2011-06-11 19:24:28	NaN	NaN	NaN
2011-06-11 19:41:25	NaN	NaN	NaN
2011-06-11 19:57:39	NaN	NaN	NaN
2011-06-11 20:14:14	NaN	NaN	NaN

```
In [37]: sd[['behavior', "F719"),
              ('latitude', "F719"),
              ('longitude', "F719")]].dropna()[[:5]]
```

```
Out[37]:
```

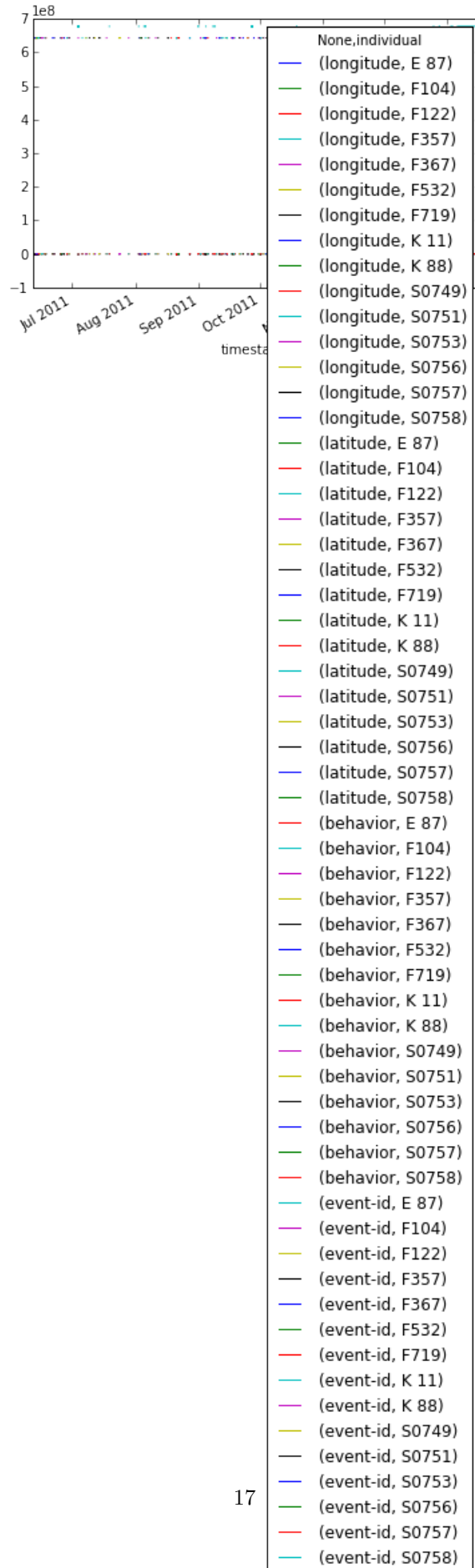
	behavior	latitude	longitude
individual	F719	F719	F719
timestamp			
2011-06-14 19:59:29	0.234145	43.937832	-59.826447
2011-06-14 20:32:38	-1.316450	43.931831	-59.829632
2011-06-14 20:48:07	0.094999	43.936577	-59.813534
2011-06-14 21:18:03	-0.078049	43.942875	-59.812885
2011-06-14 21:33:03	0.808133	43.941067	-59.808041

2.3 Simple Plotting

- Plot the data from

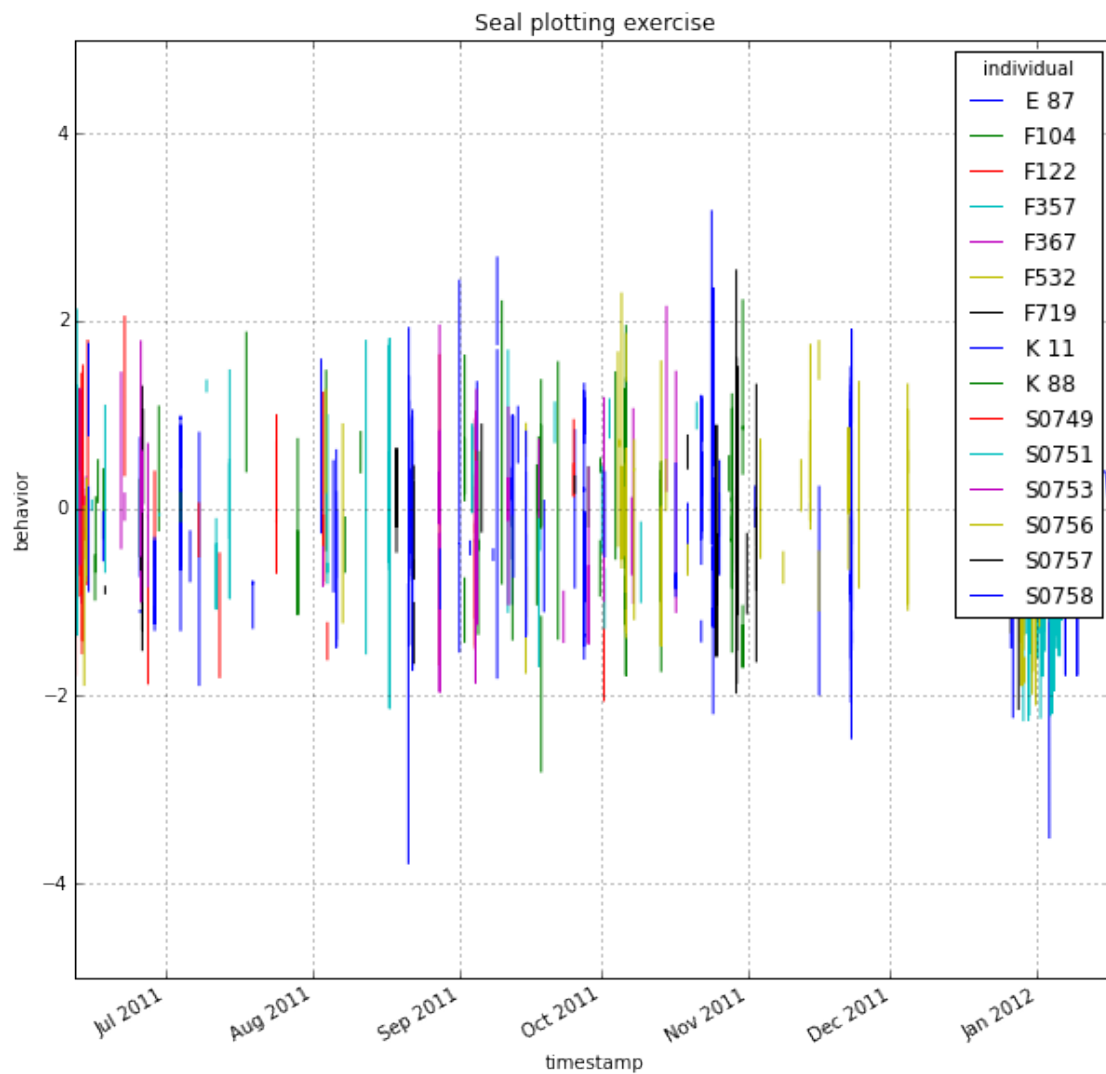
```
In [42]: sd.plot()
```

```
Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x10b053e10>
```

2.3.1 Plot the behavior of all seals

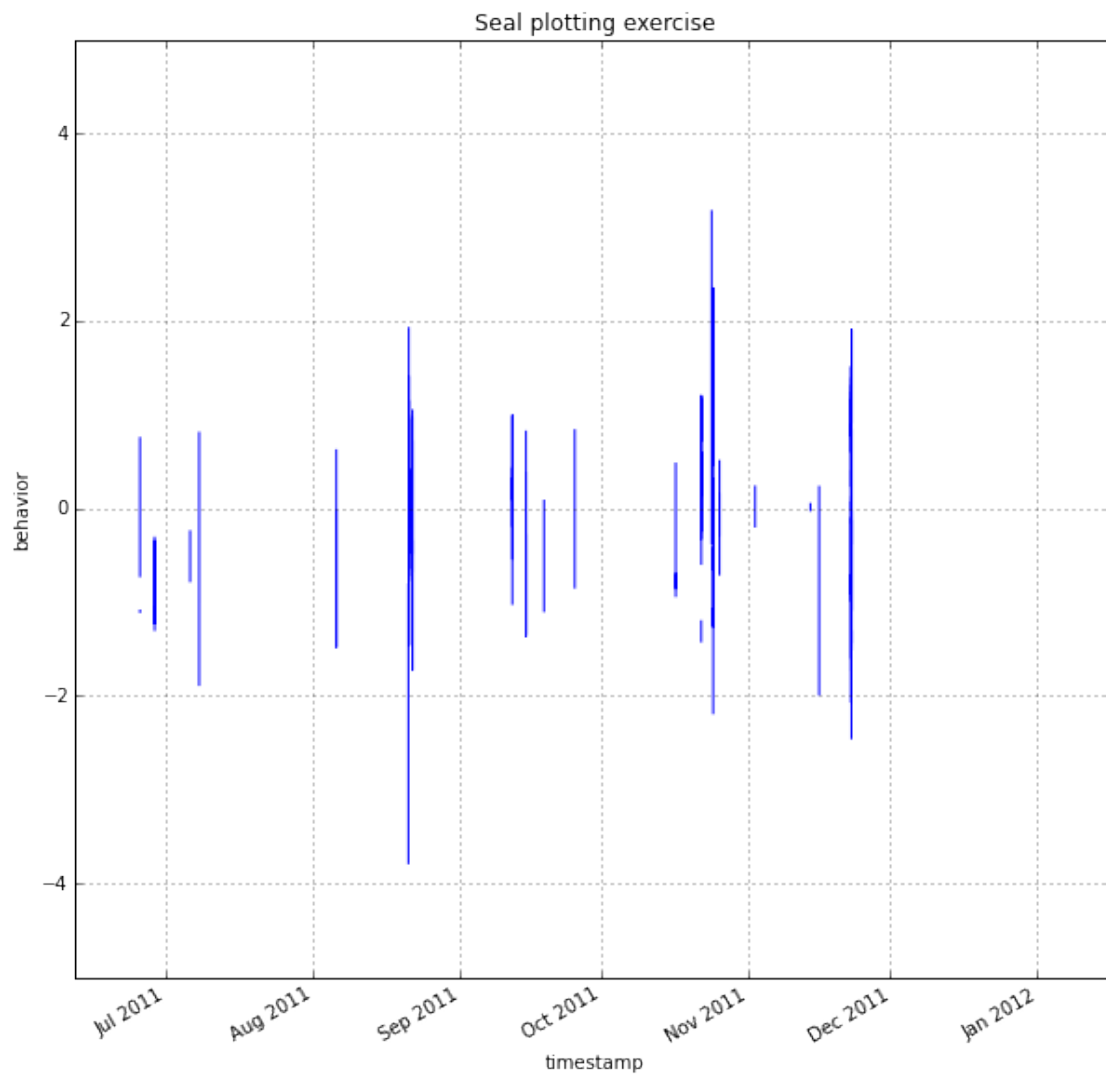
```
In [40]: sd['behavior'].plot(figsize=(10, 10))
plt.ylabel('behavior')
plt.title('Seal plotting exercise')
plt.savefig('seal_behavior.png')
plt.grid()
plt.show()
```



2.3.2 Plot the behavior of seal K 11

```
In [41]: sd(['behavior', 'K 11']).plot(figsize=(10, 10))
plt.ylabel('behavior')
plt.title('Seal plotting exercise')
```

```
plt.savefig('seal_behavior.png')
plt.grid()
plt.show()
```



2.3.3 Importing the image into the Markdown

In []: