Example of SFrame

October 28, 2015

1 Fire up GraphLab Create

```
In [1]: import graphlab
```

2 load a tabular data set

```
In [2]: sf = graphlab.SFrame('people-example.csv')

[INFO] This non-commercial license of GraphLab Create is assigned to chao@ou.eduand will expire on Octol

[INFO] Start server at: ipc:///tmp/graphlab_server-15413 - Server binary: /home/chao/anaconda/envs/dato-
[INFO] GraphLab Server Version: 1.6.1

PROGRESS: Finished parsing file /home/chao/Documents/Python/people-example.csv
PROGRESS: Parsing completed. Parsed 7 lines in 0.011037 secs.
------------------------------------------------------
Inferred types from first line of file as
column_type_hints=[str,str,str,int]
If parsing fails due to incorrect types, you can correct
the inferred type list above and pass it to read_csv in
the column_type_hints argument
------------------------------------------------------
PROGRESS: Finished parsing file /home/chao/Documents/Python/people-example.csv
PROGRESS: Parsing completed. Parsed 7 lines in 0.009043 secs.
```

3 SFrame basics

```
In [3]: sf # we can view first few lines of table

Out[3]: Columns:
                First Name        str
                Last Name        str
                Country        str
                age        int

        Rows: 7

        Data:
        +------------+-----------+---------------+-----+
        | First Name | Last Name |    Country    | age |
        +------------+-----------+---------------+-----+
        |    Bob     |   Smith   | United States |  24 |
```

```
|   Alice    |  Williams |     Canada     |  23 |
|  Malcolm   |    Jone   |    England     |  22 |
|   Felix    |   Brown   |      USA       |  23 |
|    Alex    |   Cooper  |     Poland     |  23 |
|    Tod     |  Campbell | United States  |  22 |
|   Derek    |    Ward   |  Switzerland   |  25 |
+------------+-----------+----------------+-----+
[7 rows x 4 columns]
```

In [4]:
```
sf.head()
```

Out[4]:
```
Columns:
        First Name        str
        Last Name        str
        Country        str
        age        int

Rows: 7

Data:
+------------+-----------+----------------+-----+
| First Name | Last Name |    Country     | age |
+------------+-----------+----------------+-----+
|    Bob     |   Smith   | United States  |  24 |
|   Alice    |  Williams |     Canada     |  23 |
|  Malcolm   |    Jone   |    England     |  22 |
|   Felix    |   Brown   |      USA       |  23 |
|    Alex    |   Cooper  |     Poland     |  23 |
|    Tod     |  Campbell | United States  |  22 |
|   Derek    |    Ward   |  Switzerland   |  25 |
+------------+-----------+----------------+-----+
[7 rows x 4 columns]
```

In [5]:
```
sf.tail()
```

Out[5]:
```
Columns:
        First Name        str
        Last Name        str
        Country        str
        age        int

Rows: 7

Data:
+------------+-----------+----------------+-----+
| First Name | Last Name |    Country     | age |
+------------+-----------+----------------+-----+
|    Bob     |   Smith   | United States  |  24 |
|   Alice    |  Williams |     Canada     |  23 |
|  Malcolm   |    Jone   |    England     |  22 |
|   Felix    |   Brown   |      USA       |  23 |
|    Alex    |   Cooper  |     Poland     |  23 |
|    Tod     |  Campbell | United States  |  22 |
|   Derek    |    Ward   |  Switzerland   |  25 |
+------------+-----------+----------------+-----+
[7 rows x 4 columns]
```

4 GraphLab Canvas

```
In [6]: # take any data structure in GraphLab Create
        sf.show()
```

```
Canvas is accessible via web browser at the URL: http://localhost:59608/index.html
Opening Canvas in default web browser.
```

```
In [7]: graphlab.canvas.set_target('ipynb')
```

```
In [8]: sf['age'].show(view = 'Categorical')
```

5 Insepect columns of dataset

```
In [9]: sf['Country']
```

```
Out[9]: dtype: str
        Rows: 7
        ['United States', 'Canada', 'England', 'USA', 'Poland', 'United States', 'Switzerland']
```

```
In [10]: sf['age']
```

```
Out[10]: dtype: int
         Rows: 7
         [24, 23, 22, 23, 23, 22, 25]
```

```
In [11]: sf['age'].mean()
```

```
Out[11]: 23.142857142857146
```

```
In [12]: sf['age'].max()
```

```
Out[12]: 25
```

6 Create new columns in our SFrame

```
In [13]: sf
```

```
Out[13]: Columns:
                First Name        str
                Last Name        str
                Country        str
                age        int

         Rows: 7

         Data:
         +------------+-----------+---------------+-----+
         | First Name | Last Name |    Country    | age |
         +------------+-----------+---------------+-----+
         |    Bob     |   Smith   | United States |  24 |
         |   Alice    |  Williams |     Canada    |  23 |
         |  Malcolm   |    Jone   |    England    |  22 |
         |   Felix    |   Brown   |      USA      |  23 |
         |    Alex    |   Cooper  |     Poland    |  23 |
         |    Tod     |  Campbell | United States |  22 |
         |   Derek    |    Ward   |  Switzerland  |  25 |
         +------------+-----------+---------------+-----+
         [7 rows x 4 columns]
```

```
In [14]: sf['Full Name'] = sf['First Name'] + ' ' + sf['Last Name']

In [15]: sf

Out[15]: Columns:
         First Name        str
         Last Name        str
         Country        str
         age        int
         Full Name        str

         Rows: 7

         Data:
         +------------+-----------+---------------+-----+----------------+
         | First Name | Last Name |    Country    | age |   Full Name    |
         +------------+-----------+---------------+-----+----------------+
         |    Bob     |   Smith   | United States |  24 |   Bob Smith    |
         |   Alice    |  Williams |     Canada    |  23 | Alice Williams |
         |  Malcolm   |    Jone   |    England    |  22 |  Malcolm Jone  |
         |   Felix    |   Brown   |      USA      |  23 |  Felix Brown   |
         |    Alex    |   Cooper  |     Poland    |  23 |  Alex Cooper   |
         |    Tod     |  Campbell | United States |  22 |  Tod Campbell  |
         |   Derek    |    Ward   |  Switzerland  |  25 |   Derek Ward   |
         +------------+-----------+---------------+-----+----------------+
         [7 rows x 5 columns]

In [16]: sf['age'] +2

Out[16]: dtype: int
         Rows: 7
         [26, 25, 24, 25, 25, 24, 27]

In [17]: sf

Out[17]: Columns:
         First Name        str
         Last Name        str
         Country        str
         age        int
         Full Name        str

         Rows: 7

         Data:
         +------------+-----------+---------------+-----+----------------+
         | First Name | Last Name |    Country    | age |   Full Name    |
         +------------+-----------+---------------+-----+----------------+
         |    Bob     |   Smith   | United States |  24 |   Bob Smith    |
         |   Alice    |  Williams |     Canada    |  23 | Alice Williams |
         |  Malcolm   |    Jone   |    England    |  22 |  Malcolm Jone  |
         |   Felix    |   Brown   |      USA      |  23 |  Felix Brown   |
         |    Alex    |   Cooper  |     Poland    |  23 |  Alex Cooper   |
         |    Tod     |  Campbell | United States |  22 |  Tod Campbell  |
         |   Derek    |    Ward   |  Switzerland  |  25 |   Derek Ward   |
         +------------+-----------+---------------+-----+----------------+
         [7 rows x 5 columns]
```

7 Use the apply function to do a advanced tranformation of our data

```
In [18]: sf['Country']
```

```
Out[18]: dtype: str
         Rows: 7
         ['United States', 'Canada', 'England', 'USA', 'Poland', 'United States', 'Switzerland']
```

```
In [19]: sf['Country'].show()
```

```
In [20]: def transfrom_country(country):
             if country == 'USA':
                 return 'United States'
             else:
                 return country
```

```
In [21]: transfrom_country('USA')
```

```
Out[21]: 'United States'
```

```
In [22]: sf['Country'].apply(transfrom_country)
```

```
Out[22]: dtype: str
         Rows: 7
         ['United States', 'Canada', 'England', 'United States', 'Poland', 'United States', 'Switzerlan
```

```
In [23]: sf['Country']
```

```
Out[23]: dtype: str
         Rows: 7
         ['United States', 'Canada', 'England', 'USA', 'Poland', 'United States', 'Switzerland']
```

```
In [24]: sf
```

```
Out[24]: Columns:
                 First Name        str
                 Last Name        str
                 Country        str
                 age        int
                 Full Name        str

         Rows: 7

         Data:
         +------------+-----------+---------------+-----+----------------+
         | First Name | Last Name |    Country    | age |   Full Name    |
         +------------+-----------+---------------+-----+----------------+
         |    Bob     |   Smith   | United States |  24 |   Bob Smith    |
         |   Alice    |  Williams |     Canada    |  23 | Alice Williams |
         |  Malcolm   |    Jone   |    England    |  22 |  Malcolm Jone  |
         |   Felix    |   Brown   |      USA      |  23 |  Felix Brown   |
         |    Alex    |   Cooper  |     Poland    |  23 |  Alex Cooper   |
         |    Tod     |  Campbell | United States |  22 |  Tod Campbell  |
         |   Derek    |    Ward   |  Switzerland  |  25 |   Derek Ward   |
         +------------+-----------+---------------+-----+----------------+
         [7 rows x 5 columns]
```

```
In [25]: sf['Country']=sf['Country'].apply(transfrom_country)

In [26]: sf

Out[26]: Columns:
                First Name        str
                Last Name        str
                Country        str
                age        int
                Full Name        str

         Rows: 7

         Data:
         +------------+-----------+---------------+-----+----------------+
         | First Name | Last Name |    Country    | age |   Full Name    |
         +------------+-----------+---------------+-----+----------------+
         |    Bob     |   Smith   | United States |  24 |   Bob Smith    |
         |   Alice    |  Williams |     Canada    |  23 | Alice Williams |
         |  Malcolm   |    Jone   |    England    |  22 |  Malcolm Jone  |
         |   Felix    |   Brown   | United States |  23 |  Felix Brown   |
         |    Alex    |   Cooper  |     Poland    |  23 |  Alex Cooper   |
         |    Tod     |  Campbell | United States |  22 |  Tod Campbell  |
         |   Derek    |    Ward   |  Switzerland  |  25 |   Derek Ward   |
         +------------+-----------+---------------+-----+----------------+
         [7 rows x 5 columns]

In [ ]:
```