DSFA

Spring 2018

Lecture 6

Census

Announcements

- Homework 2 due Thursday 2/8
- Lab 3 will be released tomorrow

Tables Review

Table Structure

- A Table is a sequence of labeled columns
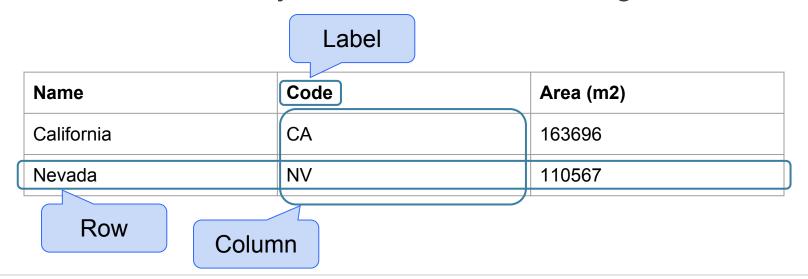- Labels are strings
- Columns are arrays, all with the same length

Label

Name	Code	Area (m2)
California	CA	163696
Nevada	NV	110567

Row

Column

Table Methods

- Creating and extending tables:
  - `Table().with_columns` and `Table.read_table`
- Finding the size: `t.num_rows` and `t.num_columns`
- Referring to columns: labels, relabeling, and indices
  - `t.labels` and `t.relabeled`; column indices start at 0
- Accessing data in a column
  - `t.column` takes a label or index and returns an array
- Using array methods to work with data in columns
  - `a.item(row_index)` returns a value in an array
  - `a.sum()`, `a.min()`, `a.max()` or `sum(a)`, `min(a)`, `max(a)`
- Creating new tables containing some of the original columns:
  - `select`, `drop`

Manipulating Rows

- `t.`**`sort`**`(column)` sorts the rows in increasing order
- `t.`**`take`**`(row_numbers)` keeps the numbered rows
  - Each row has an index, starting at 0
- `t.`**`where`**`(column, are.condition)` keeps all rows for which a column's value satisfies a condition
- `t.`**`where`**`(column, value)` keeps all rows for which a column's value equals some particular value
- `t.`**`with_row`** makes a new table that has another row

(Demo)

Discussion Questions

The table **nba** has columns **NAME**, **POSITION**, and **SALARY**.

a) Create an array containing the names of all point guards (**PG**) who make more than $15M/year

```
nba.where(1, 'PG').where(2, are.above(15)).column(0)
```

b) After evaluating these two expressions in order, what's the result of the second one?

```
nba.with_row(['Samosa', 'Mascot', 100])
nba.where('NAME', are.containing('Samo'))
```

Census Data

The Decennial Census

- Every ten years, the Census Bureau counts how many people there are in the U.S.

- In between censuses, the Bureau estimates how many people there are each year.

- Article 1, Section 2 of the Constitution:
  - "Representatives and direct Taxes shall be apportioned among the several States … according to their respective Numbers …"

Analyzing Census Data

Leads to the discovery of interesting features and trends in the population

(Demo)

Census Table Description

- Values have column-dependent interpretations
  - The SEX column: 1 is *Male*, 2 is *Female*
  - The POPESTIMATE2010 column: *7/1/2010 estimate*
- In this table, some rows are sums of other rows
  - The SEX column: 0 is *Total* (of *Male* + *Female*)
  - The AGE column: 999 is *Total* of all ages
- Numeric codes are often used for storage efficiency
- Values in a column have the same type, but are not necessarily comparable (AGE 12 vs AGE 999)