DSC 10, Spring 2018
Lecture 5

Tables II

sites.google.com/eng.ucsd.edu/dsc-10-spring-2018

Credit: Anindita Adhikari and John DeNero

RACE TO THE POLLS

Vote on TRITONLINK April 9-13

Announcements

- Lab 2 due tonight 11:59pm
- Please fill out survey about last night’s guest lecture
- Piazza use:
  - Ok to share your code in a private post to instructors
  - Not ok to share your code in a public post
- In-person help:
  - Go to the lab (CSE B230) during tutor office hours
  - Visit autograder.ucsd.edu
  - Submit a ticket with your location and issue to let the tutor know you need help

Sort

Sorting Tables

Tables are ordered collections of rows

- The `sort` method creates a new table with the same rows in a different order (the original table is unaffected)
- The `show` method displays the first rows of a table

(Demo)

Discussion Question

To create a table of the highest-paid players in each position:

```
nba.sort(3, descending=True).sort(1, distinct=True)
```

Which code creates a table of the lowest-paid players in each position?

A. `nba.sort(3, descending=True).sort(1, distinct=False)`

B. `nba.sort(3, descending=False).sort(1, distinct=True)`

C. `nba.sort(3, descending=False).sort(1, distinct=False)`

D. `nba.sort(3, descending=True).sort(1, distinct=True)`

Lists

Lists are Generic Sequences

A list is a sequence of values (just like an array), but the values can all have different types

```
[2+3, 'four', Table().with_column('K', [3, 4])]
```

If you create a table column from a list, it will be converted to an array automatically

(Demo)

Take

Take Rows, Select Columns

The `select` method returns a table with only some columns

The `take` method returns a table with only some rows

Take Rows, Select Columns

The `select` method returns a table with only some columns

The `take` method returns a table with only some rows

- Rows are numbered, starting at 0
- Taking a single number returns a one-row table
- Taking a list of numbers returns a table as well

(Demo)

Where

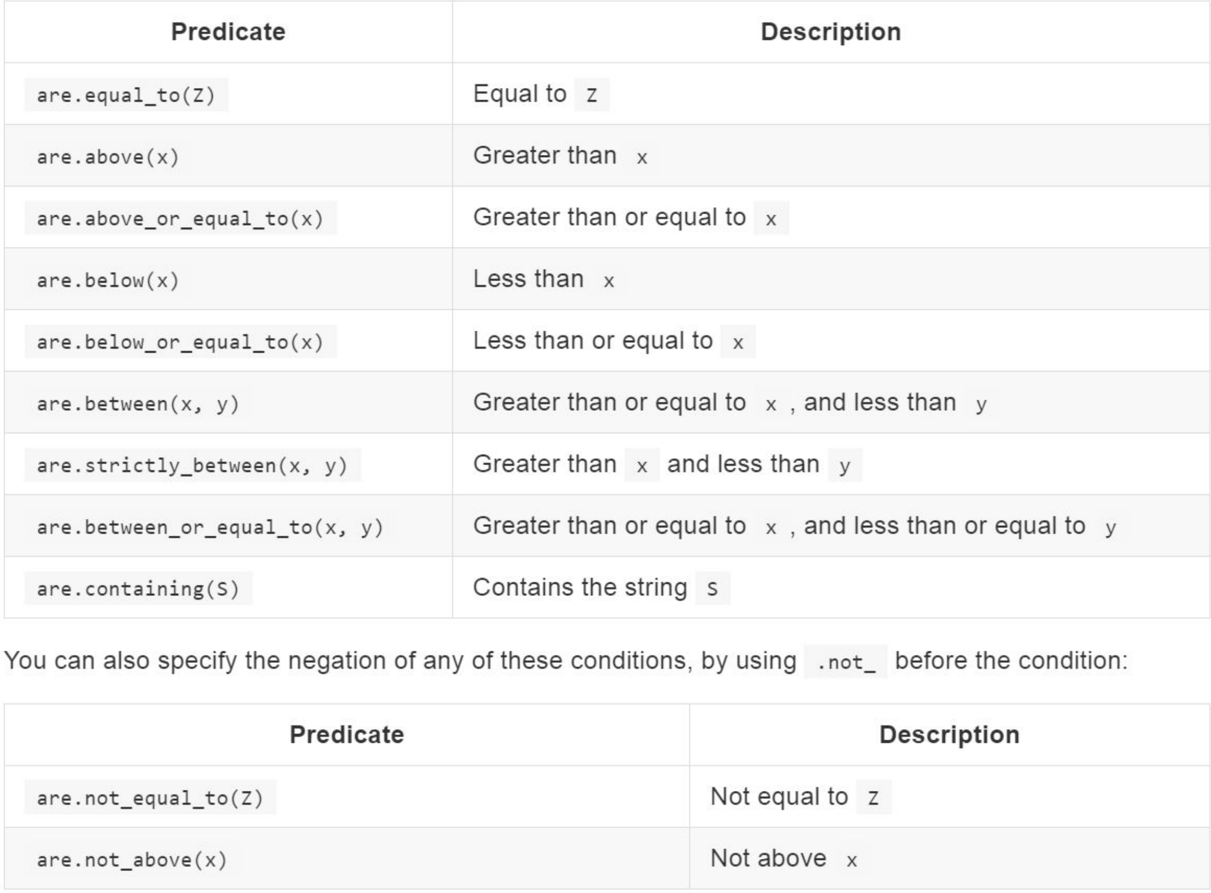
The Where Method

The `where` method specifies a column and a condition

It returns a new table with all rows satisfying the condition

(Demo)

Some Conditions



Predicate	Description
`are.equal_to(Z)`	Equal to `Z`
`are.above(x)`	Greater than `x`
`are.above_or_equal_to(x)`	Greater than or equal to `x`
`are.below(x)`	Less than `x`
`are.below_or_equal_to(x)`	Less than or equal to `x`
`are.between(x, y)`	Greater than or equal to `x` , and less than `y`
`are.strictly_between(x, y)`	Greater than `x` and less than `y`
`are.between_or_equal_to(x, y)`	Greater than or equal to `x` , and less than or equal to `y`
`are.containing(S)`	Contains the string `S`

You can also specify the negation of any of these conditions, by using `.not_` before the condition:

Predicate	Description
`are.not_equal_to(Z)`	Not equal to `Z`
`are.not_above(x)`	Not above `x`

Discussion Question

The table `nba` has columns `PLAYER`, `POSITION`, `TEAM`, `SALARY`.

Order the snippets of code to calculate the total salary of all small forwards (`SF`)

A. `nba` `.column(3)` `.sum()` `.where(1,'SF')`

B. `nba` `.where(1,'SF')` `.sum()` `.column(3)`

C. `nba` `.column(3)` `.where(1,'SF')` `.sum()`

D. `nba` `.where(1,'SF')` `.column(3)` `.sum()`

Discussion Question

The table `nba` has columns `PLAYER`, `POSITION`, `TEAM`, `SALARY`.

```
nba.where(1,'SF').column(3).sum()/nba.where(1,'SF').num_rows
```

What does this code compute?

(Demo)

Discussion Question

The table `nba` has columns `PLAYER`, `POSITION`, `TEAM`, `SALARY`.

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

Discussion Question

The table **nba** has columns **PLAYER**, **POSITION**, **TEAM**, **SALARY**.

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

```
nba.where(1, 'PG').where(3, are.above(15)).column(0)
```

(Demo)

Discussion Question

The table `nba` has columns `PLAYER`, `POSITION`, `TEAM`, `SALARY`.

What is the output when we execute a cell containing these two lines of code?

```
nba.with_row(['Jazz Bear','Mascot','Utah Jazz',100])
nba.where('PLAYER', are.containing('Bear'))
```

A. A table with one row for Jazz Bear
B. An empty table with no rows
C. An error message

(Demo)

Summary of Manipulating Rows

- `t.sort(column)`
  - sorts the rows in increasing order
- `t.take(row_numbers)`
  - keeps only specified rows (row numbers start 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 containing a certain value in a column

Practice

The table `menu` has a row for each item on a restaurant's menu. The columns are `Item` and `Price`, in that order. One of the menu items is `Cheeseburger.`

Write one line of code that produces the same table without a row for `Cheeseburger.`

Practice

The table `menu` has a row for each item on a restaurant's menu. The columns are `Item` and `Price`, in that order. One of the menu items is `Cheeseburger.`

Write one line of code that produces the same table without a row for `Cheeseburger.`

```
menu.where('Item', are.not_equal_to('Cheeseburger'))
```

Practice

The table `menu` has a row for each item on a restaurant’s menu. The columns are `Item` and `Price`, in that order. One of the menu items is `Cheeseburger.`

Which line of code finds the number of items on the menu at this restaurant?

A. `menu.num_rows`
B. `menu.column(0).num_rows`
C. `menu.column(0).length`
D. `menu.column(1).size`
E. `More than one of the above`

Practice

The table `menu` has a row for each item on a restaurant's menu. The columns are `Item` and `Price`, in that order. One of the menu items is `Cheeseburger.`

Write one line of code that evaluates to

a) the name of a menu item that has the lowest possible price.

b) **Challenge**: a table containing the name of **all** menu items that have the lowest possible price.

Practice

The table **`menu`** has a row for each item on a restaurant's menu. The columns are **`Item`** and **`Price`**, in that order. One of the menu items is **`Cheeseburger.`**

Write one line of code that evaluates to

a) the name of a menu item that has the lowest possible price.

```
menu.sort(‘Price’).column(0).item(0)
```

b) **Challenge**: a table containing the names of **all** menu items that have the lowest possible price.

Practice

The table `menu` has a row for each item on a restaurant's menu. The columns are `Item` and `Price`, in that order. One of the menu items is `Cheeseburger.`

Write one line of code that evaluates to

a) the name of a menu item that has the lowest possible price.

```
menu.sort(‘Price’).column(0).item(0)
```

b) **Challenge**: a table containing the names of **all** menu items that have the lowest possible price.

```
menu.sort(‘Price’).where(‘Price’,
menu.sort(‘Price’).column(‘Price).item(0)).select(‘Item’)
```

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.

Why estimate each year?

Why estimate 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 …”

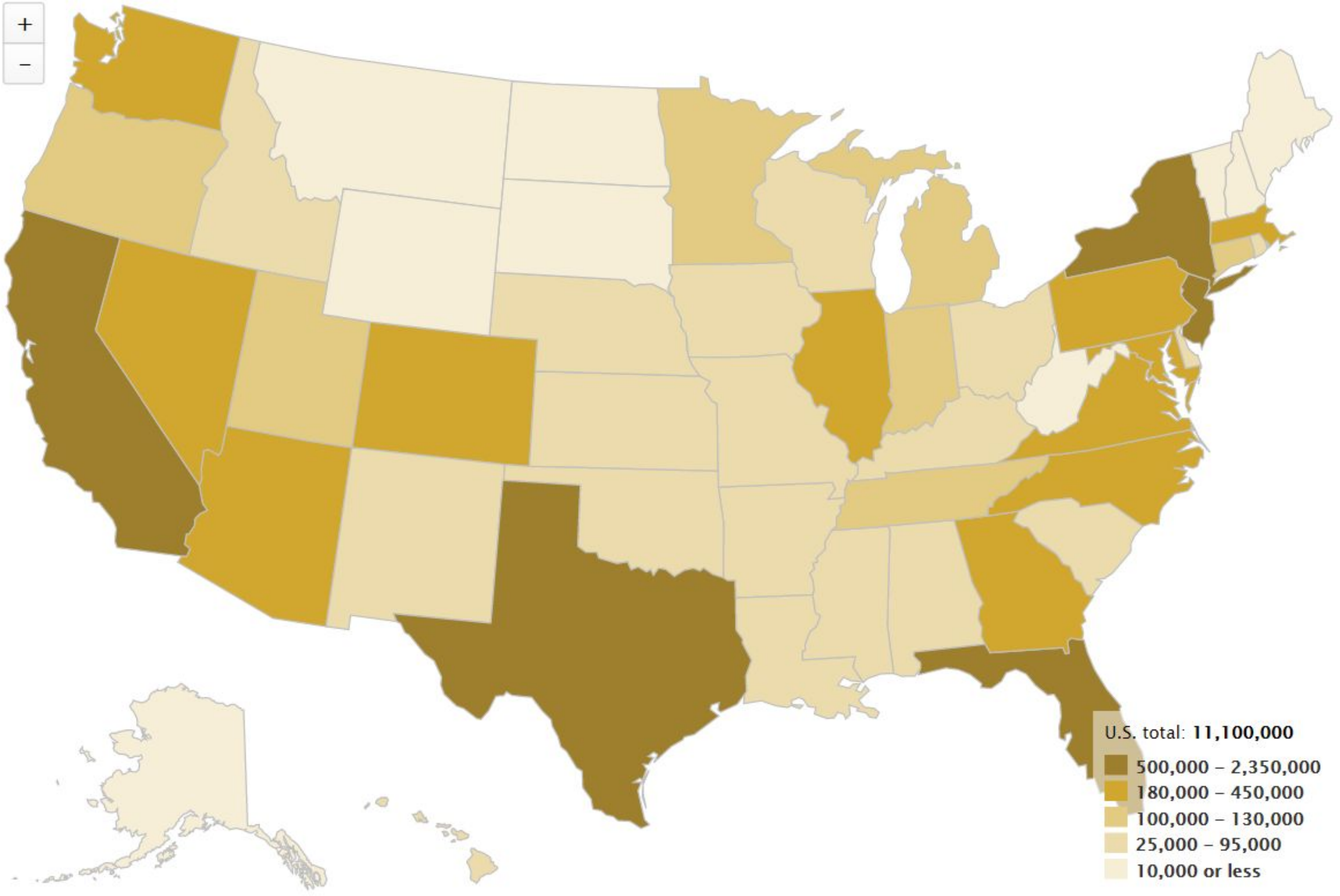
Why estimate each year?

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

Which of these states would be most likely to want to adjust the census to correct undercount?

A. Hawaii
B. Wyoming
C. Texas
D. Vermont
E. New York

Estimated unauthorized immigrant population, by state, 2014



Pew Research Center

Census Table Description

- Interpretation of values in table depend on columns
  - The SEX column: 1 is *Male*, 2 is *Female*
  - The POPESTIMATE2010 column: *7/1/2010 estimate*

http://www2.census.gov/programs-surveys/popest/datasets/2010-2015/national/asrh/nc-est2015-agesex-res.pdf

Census Table Description

- Interpretation of values in table depend on columns
  - 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

http://www2.census.gov/programs-surveys/popest/datasets/2010-2015/national/asrh/nc-est2015-agesex-res.pdf

Census Table Description

- Interpretation of values in table depend on columns
    - 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

http://www2.census.gov/programs-surveys/popest/datasets/2010-2015/national/asrh/nc-est2015-agesex-res.pdf

Census Table Description

- Interpretation of values in table depend on columns
  - 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)

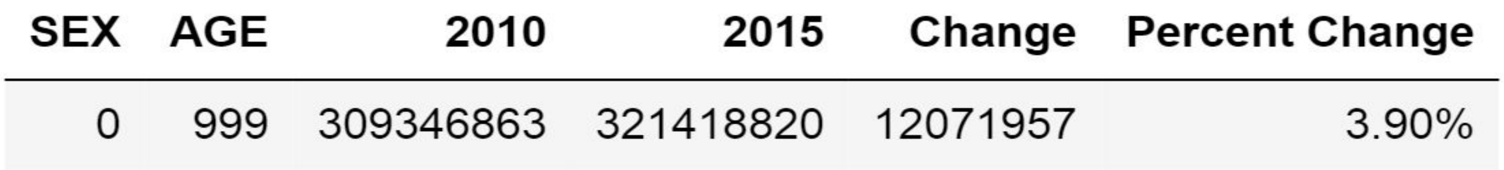
http://www2.census.gov/programs-surveys/popest/datasets/2010-2015/national/asrh/nc-est2015-agesex-res.pdf

Analyzing Census Data

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

(Demo)

Discussion Question



SEX	AGE	2010	2015	Change	Percent Change
0	999	309346863	321418820	12071957	3.90%

What does this code calculate?

```
(321418820/309346863) ** (1/5) - 1
```

A. The ratio of the population in 2015 to the population in 2010.
B. The percentage by which the population changed from 2010 to 2015.
C. The annual growth rate for the population from 2010 to 2015.
D. This code doesn’t compute a meaningful value.

(Demo)