DATA 1202

Spring 2024

Lecture 5

Building Tables

Announcements

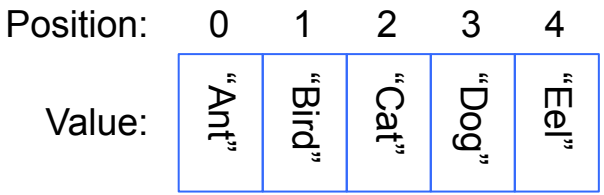
- **HW 3** is due tonight
- **Lab 3** is due on Friday

Weekly Goals

- Monday
  - Data Types
- Wednesday (Today)
  - Building Tables
  - Working with Census data

Arrays

An array contains a sequence of values



- **All elements of an array should have the same type**
- Arithmetic is applied to each element individually
- Adding two arrays adds the corresponding elements (but the arrays must be the same length!)
- A column of a table is an array

(Demo)

Columns

Columns are Arrays

A table of a column is an array of values (one per row).

For a table `t`, use `t.column(label)` or `t.column(index)`

For an array `s`:

- `s.item(index)` gives the value at an index (starting at 0)
- Two ways to aggregate the values in an array:
  - `np.mean(s)`, `np.sum(s)`, `np.max(s)`, `np.min(s)`
  - `s.mean()`, `s.sum()`, `s.max()`, `s.min()`

(Demo)

Ranges

(Demo)

Ranges

A range is an array of consecutive numbers

- **`np.arange(end)`**:
  An array of increasing integers from 0 up to **`end`**
- **`np.arange(start, end)`**:
  An array of increasing integers from **`start`** up to **`end`**
- **`np.arange(start, end, step)`**:
  A range with **`step`** between consecutive values

The range always includes **`start`** but excludes **`end`**

Creating Tables

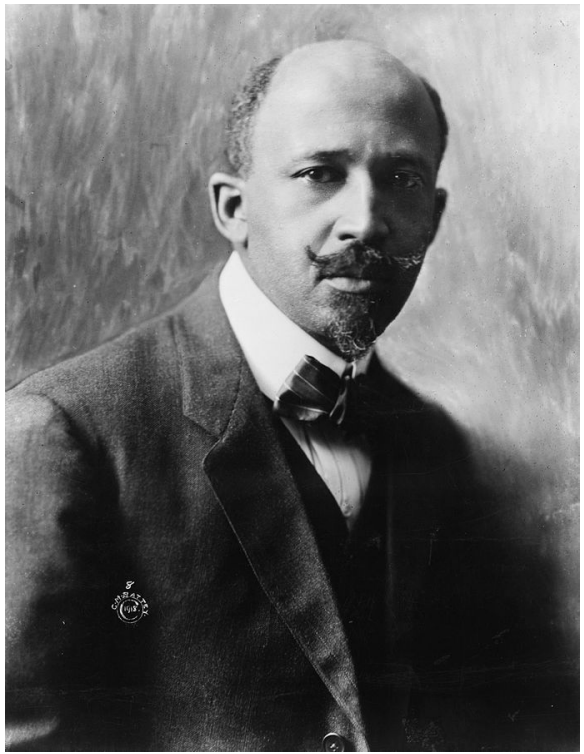
Ways to create a table

- `Table.read_table(filename)` - reads a table from a file (such as a spreadsheet)
- `Table()` - an empty table to which columns are added
- and… `select, where, sort, drop, take` all create new tables based on existing tables

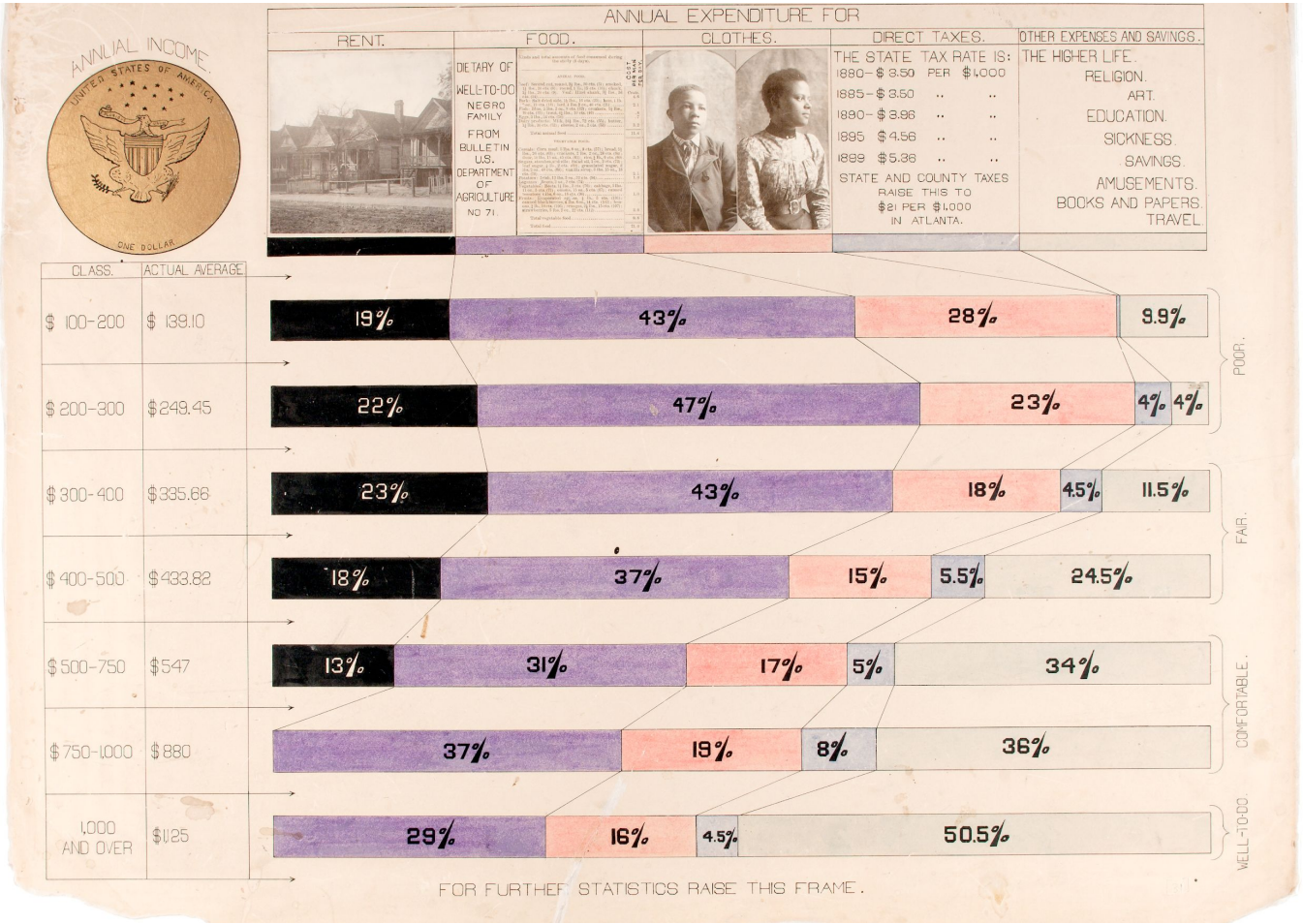
(Demo)

Example

W. E. B. Du Bois, 1868-1963



- Scholar, historian, activist, and data scientist
- A founder of the NAACP
- Made a series of visualizations for the 1900 Paris Exposition
  - Goal: educate people about the lives of Black Americans
  - Hundreds of photographs and patents
  - About 30 handmade statistical graphics (created in 3 months)
  - Now in the Library of Congress



This is a photograph of the actual graph produced by Du Bois and his graduate students. The title (omitted) contains language that was used at the time, but would not be appropriate to use today.

(Demo)

ANNUAL INCOME.

UNITED STATES OF AMERICA

ONE DOLLAR

ANNUAL EXPENDITURE FOR

RENT.	FOOD.	CLOTHES.	DIRECT TAXES.	OTHER EXPENSES AND SAVINGS.
	DIETARY OF WELL-TO-DO NEGRO FAMILY FROM BULLETIN U.S. DEPARTMENT OF AGRICULTURE NO 71. 		THE STATE TAX RATE IS: 1880–$3.50 PER $1,000 1885–$3.50 .. .. 1890–$3.96 .. .. 1895 $4.56 .. .. 1899 $5.36 .. .. STATE AND COUNTY TAXES RAISE THIS TO $21 PER $1,000 IN ATLANTA.	THE HIGHER LIFE. RELIGION. ART. EDUCATION. SICKNESS. SAVINGS. AMUSEMENTS. BOOKS AND PAPERS. TRAVEL.

CLASS.	ACTUAL AVERAGE						
$ 100-200	$ 139.10	19%	43%	28%	9.9%		POOR.
$ 200-300	$ 249.45	22%	47%	23%	4%	4%	
$ 300-400	$ 335.66	23%	43%	18%	4.5%	11.5%	FAIR.
$ 400-500	$ 433.82	18%	37%	15%	5.5%	24.5%	
$ 500-750	$ 547	13%	31%	17%	5%	34%	COMFORTABLE.
$ 750-1,000	$ 880	37%	19%	8%	36%		
1,000 AND OVER	$ 1125	29%	16%	4.5%	50.5%		WELL-TO-DO.

Discussion Question

Use the table functions we learned this week to find the income bracket (“class”) that spent the highest percentage of their income on rent.

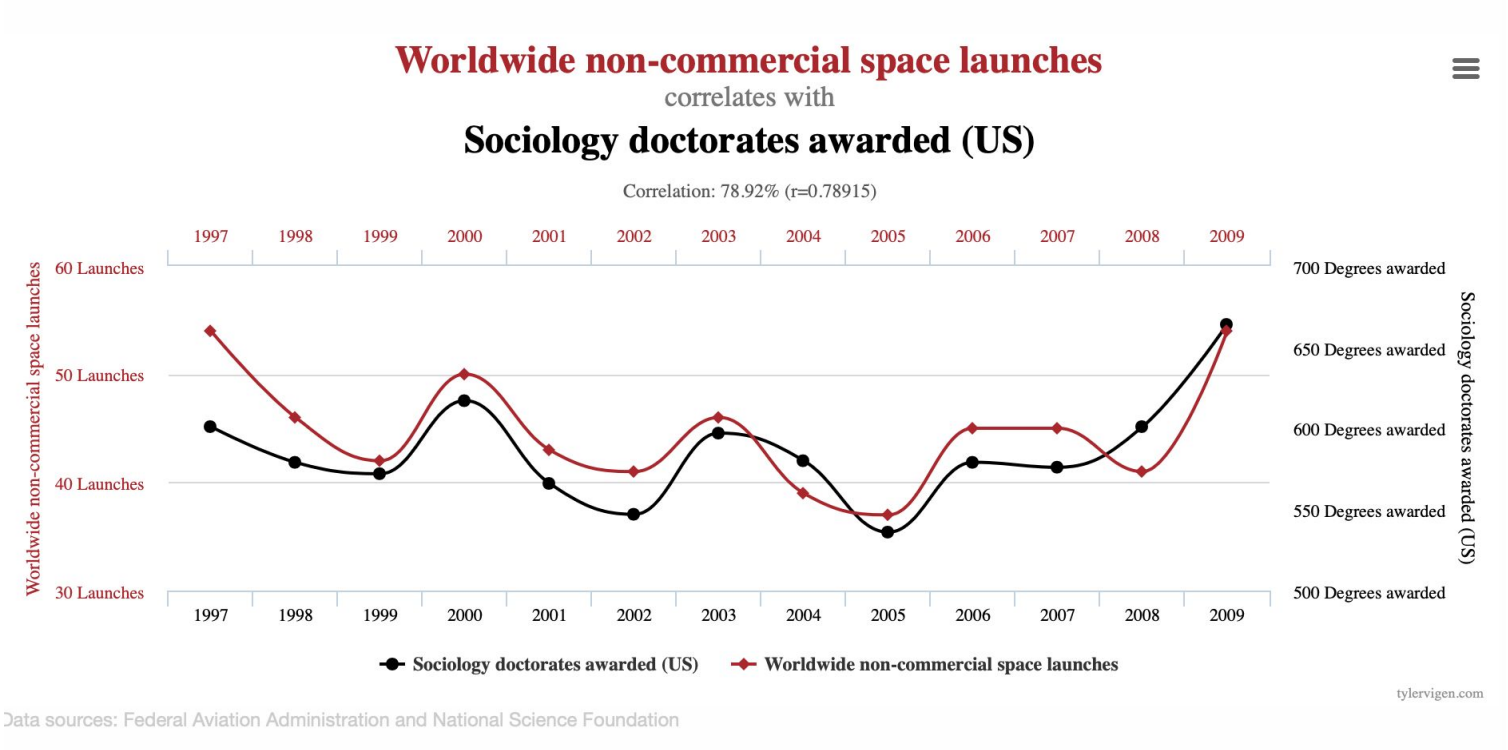
Table Methods So Far

- Creating tables: `Table().with_columns` ; `Table.read_table`
- Finding the size: `t.num_rows` and `t.num_columns`
- Listing/changing the column labels: `t.labels` and `t.relabeled`
- An array of column values: `column` takes a label or index
- A table containing some of the original columns: `select`, `drop`
- A table containing some of the original rows: `where`, `take`
- A table with the original rows sorted: `sort`

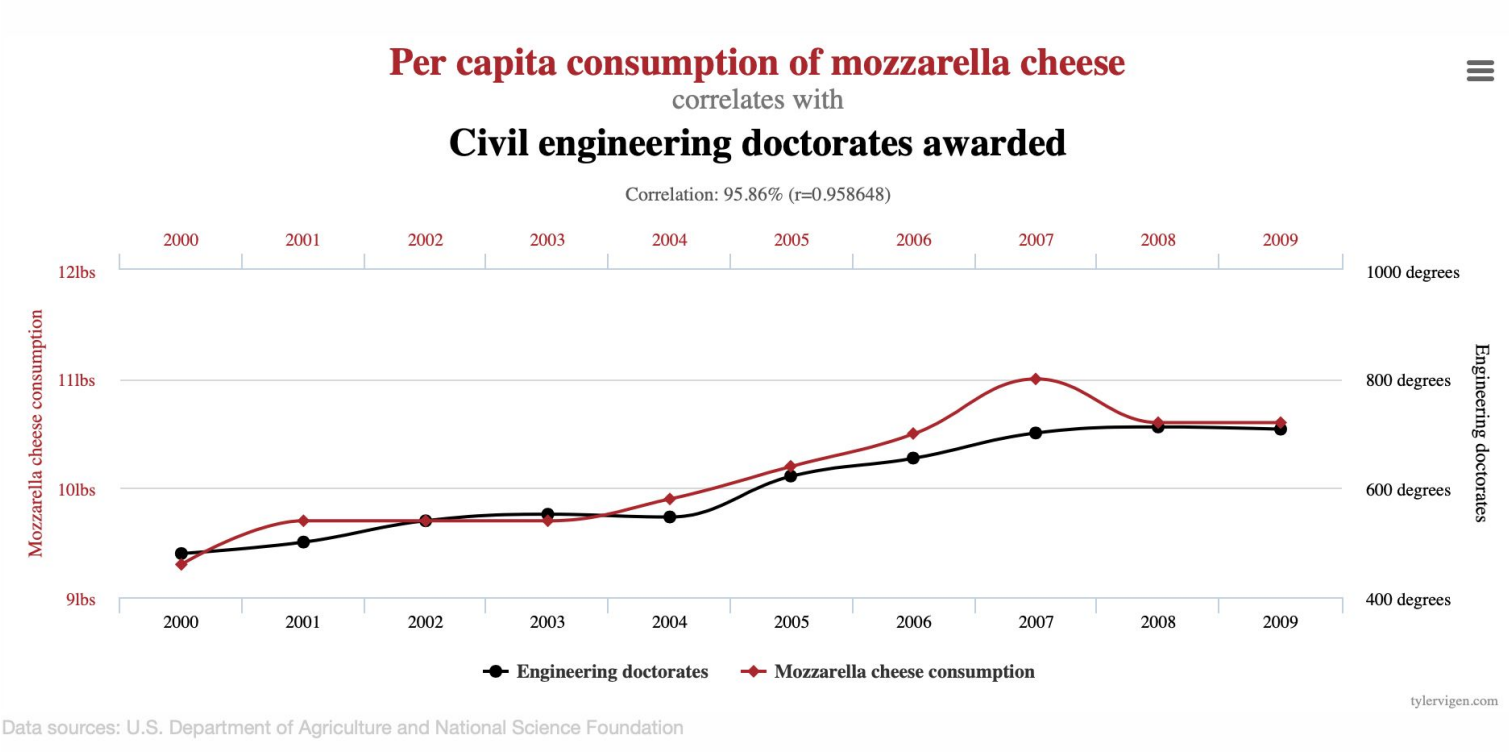
What's next? More ways of using `where` to focus on certain rows.

Association ≠ Causation

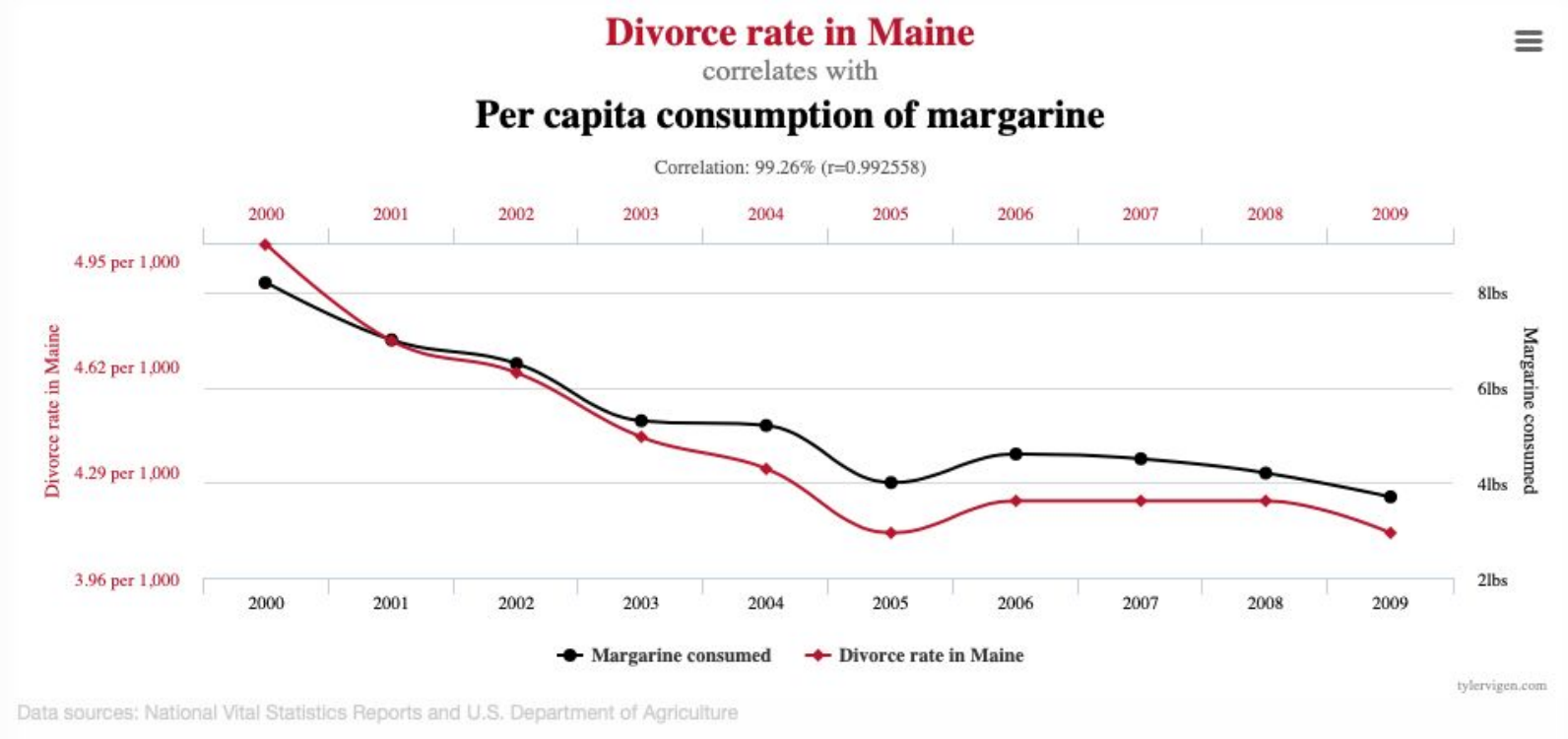
An Example 👩‍🎓 👩‍🚀


Worldwide non-commercial space launches
correlates with
Sociology doctorates awarded (US)
Correlation: 78.92% (r=0.78915)
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
60 Launches
50 Launches
40 Launches
30 Launches
Worldwide non-commercial space launches
700 Degrees awarded
650 Degrees awarded
600 Degrees awarded
550 Degrees awarded
500 Degrees awarded
Sociology doctorates awarded (US)
Sociology doctorates awarded (US)
Worldwide non-commercial space launches
tylervigen.com
Data sources: Federal Aviation Administration and National Science Foundation

One More 🧀 👷‍♀️


Per capita consumption of mozzarella cheese
correlates with
Civil engineering doctorates awarded
Correlation: 95.86% (r=0.958648)
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
12lbs
11lbs
10lbs
9lbs
Mozzarella cheese consumption
1000 degrees
800 degrees
600 degrees
400 degrees
Engineering doctorates
Engineering doctorates
Mozzarella cheese consumption
tylervigen.com
Data sources: U.S. Department of Agriculture and National Science Foundation

Another One 🧈💔


Divorce rate in Maine
correlates with
Per capita consumption of margarine
Correlation: 99.26% (r=0.992558)
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
4.95 per 1,000
4.62 per 1,000
4.29 per 1,000
3.96 per 1,000
Divorce rate in Maine
8lbs
6lbs
4lbs
2lbs
Margarine consumed
Margarine consumed
Divorce rate in Maine
tylervigen.com
Data sources: National Vital Statistics Reports and U.S. Department of Agriculture