

YData: An Introduction to Data Science

Lecture 12: Table Examples

Elena Khusainova & John Lafferty
Statistics & Data Science, Yale University
Spring 2021

Credit: data8.org



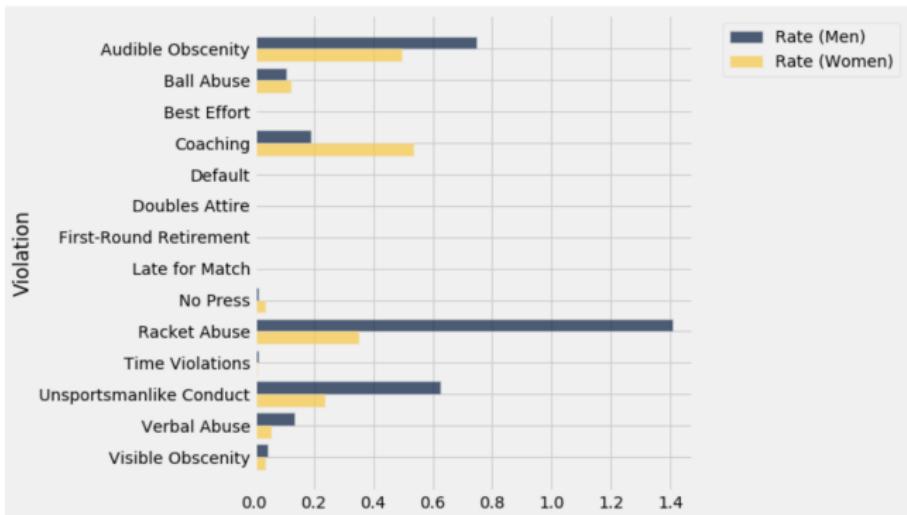
Announcements

- Project 1 checkpoint this Friday
- hw04 due this Thursday
- midterm logistics

Outline for today

- Comment on tennis data
- Maps with Covid-19 data
- Combining tables
- Sample midterm question
- Advanced where

Recall



Tennis gender bias controversy



NEWS GIRLS & WOMEN

Serena Williams Is Calling Out Sexism in Her Sport — and She's Not Alone



By Danielle Selby

SEPT. 10, 2018



38



Why Global Citizens Should Care

Around the world, women face discrimination in the workplace and their careers are often penalized because of it. Serena Williams has been a proud advocate for women's rights both on and off the court, and this is just another example of her commitment to advancing gender equality. You can take action [here](#) to help eliminate gender discrimination.

Serena Williams talks with Naomi Osaka, of Japan, after Osaka defeated Williams in the women's final of the U.S. Open tennis tournament, Sept. 8, 2018, in New York.
Adam Hunger/AP

② What is this content piece? ▾



Stay up to date

Receive alerts about the world's biggest challenges.

Tennis gender bias controversy



Nate Silver @NateSilver538



The study doesn't show that. It shows that male players are fined more, but that could be because they misbehave more. (Indeed, from watching a fair bit of tennis, the men do misbehave more). This data doesn't tell us anything about whether they're punished at greater rates.



Glenn Greenwald @ggreenwald

Replies to @ggreenwald

2/ Now, NYT just released a study of the actual data: contrary to that narrative, male tennis players are punished at far greater rates for misbehavior, especially the ones relevant to that controversy: verbal abuse, obscenity, and unsportsmanlike conduct
nytimes.com/2018/09/14/spo...

Fines by Gender at Grand Slam Tournaments

All Fines, 1998-2018	Men	Women
Racket Abuse	646	99
Audible Obscenity	344	140
Unsportsmanlike Conduct	287	67
Coaching	87	152
Ball Abuse	49	35
Verbal Abuse	62	16
Visible Obscenity	20	11

Tennis gender bias controversy

Category	Men	Women
Doubles Attire	2	1
Late for Match	1	1
First-Round Retirement	2	0
Totals	1517	535

7:08 AM · Sep 15, 2018 (i)

Heart 15.9K Comment 3.5K Share Copy link to Tweet

After tweeting at the author of the piece a couple times, Ohanian decided to focus his efforts in a more positive manner. He [donated \\$10 for every word](#) in the article to [Donors Choose](#) in order to “make sure the next generation gets access to learn basic statistics.”

[https:](https://sports.yahoo.com/serena-williams-husband-trashes-study-saying-men-penalized-women-tennis-231613134.html)

[/sports.yahoo.com/serena-williams-husband-trashes-study-saying-men-penalized-women-tennis-231613134.html](https://sports.yahoo.com/serena-williams-husband-trashes-study-saying-men-penalized-women-tennis-231613134.html)

Maps

(DEMO)

Maps

A table containing columns of latitude and longitude values can be used to generate a map of markers

_____.**map_table**(table, ...)

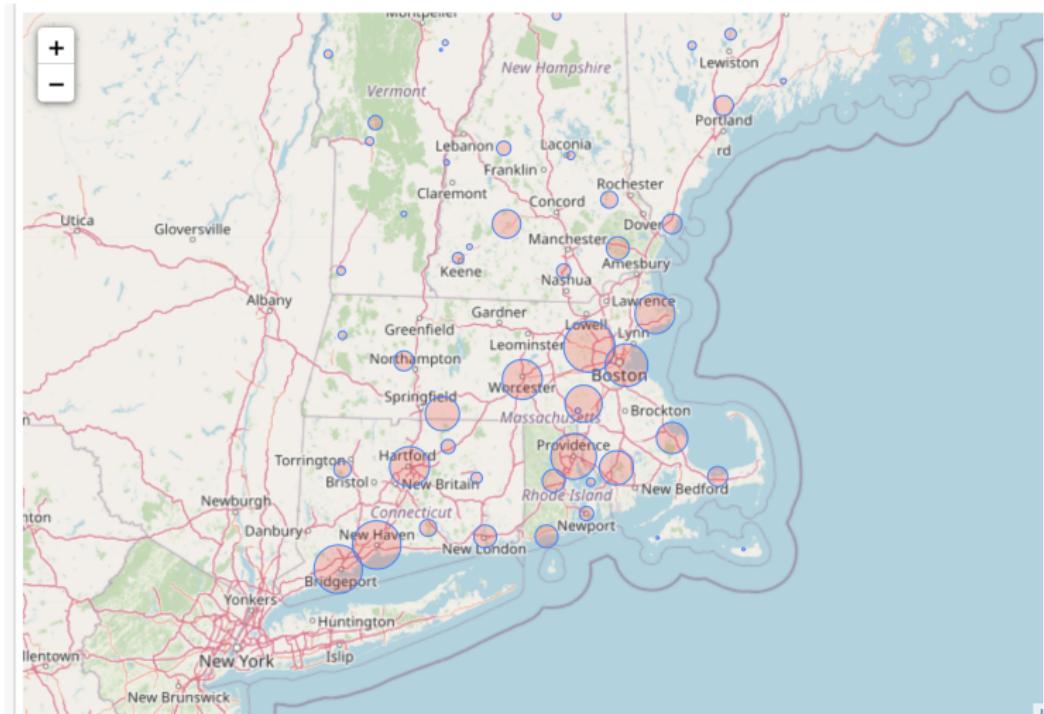
Either **Marker**
or **Circle**

Column 0: latitudes
Column 1: longitudes
Column 2: labels
Column 3: colors
Column 4: sizes

Applies to all
features:
`color='blue'`
`size=200`

Maps

We'll demonstrate with COVID-19 data



Combining Table Methods

Important Table Methods

t.select(column, ...) or t.drop(column, ...)
t.take([row, ...]), or t.exclude([row, ...])
t.sort(column, descending=False, distinct=False)
t.where(column, are.condition(...))
t.apply(function, column, ...)
t.group(column) or t.group(column, function)
t.group([column, ...]) or t.group([column, ...], function)
t.pivot(cols, rows) or t.pivot(cols, rows, vals, function)
t.join(column, other_table, other_table_column)

More documentation can be found here:

<http://data8.org/datascience/tables.html>

Discussion Question

Generate a table with one row per cafe that has the name and discounted price of its cheapest discounted drink

Drink	Cafe	Price	Coupon % off	Location
Milk Tea	Book Trader Cafe	4	5	Willoughby's
Espresso	Willoughby's	2	50	Blue State Coffee
Coffee	Willoughby's	3	25	Willoughby's
Espresso	Blue State Coffee	2	0	Book Trader Cafe

(DEMO)

Previous midterm question

- (b) (8 pt) Each row of the `trip` table from lecture describes a single bicycle rental in the San Francisco area. Durations are integers representing times in seconds. The first three rows out of 338343 appear below.

Start	End	Duration
Ferry Building	SF Caltrain	765
San Antonio Shopping Center	Mountain View City Hall	1036
Post at Kearny	2nd at South Park	307

Write a Python expression below each of the following descriptions that computes its value. The first one is provided for you. You *may* use up to two lines and introduce variables.

- The average duration of a rental.

```
total_duration = sum(trip.column(2))  
total_duration / trip.num_rows
```

(DEMO)

Advanced where

Comparison Operators

The result of a comparison expression is a `bool` value

`x = 2` `y = 3`

Assignment statements

`x > 1` `x > y` `y >= 3`

Comparison
expressions

`x == y` `x != 2` `2 < x < 5`

`t.where(array_of_bool_values)` returns a table with only the rows of `t` for which the corresponding `bool` is True.

(DEMO)