#### Learning goals:

- discuss pandas nuances
- Elections Demo
- Case Study Lab & Q&A

## Discussion 4

COGS 108 Winter 2021

# Why does my code sometimes break?

Keeping track of notebook state is subtle!

In [ ]: df income

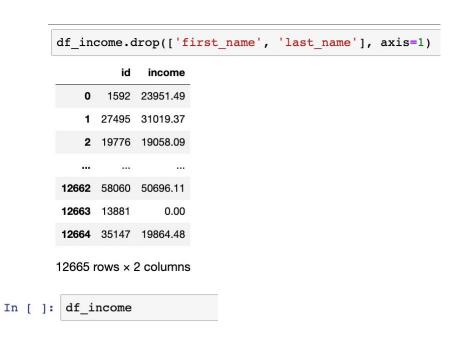
	first_name	id	income	last_name
0	Lauren	1592	23951.49	Murphy
1	Rebecca	27495	31019.37	Walls
2	Alejandra	19776	19058.09	Garcia
12662	Mark	58060	50696.11	Torres
12663	Peter	13881	0.00	Gibson
12664	Michele	35147	19864.48	Robinson

```
df_income.drop(['first_name', 'last_name'], axis=1)
              income
    0 1592 23951.49
    1 27495 31019.37
    2 19776 19058.09
12662 58060 50696.11
      13881
                0.00
12663
12664 35147 19864.48
12665 rows x 2 columns
```

	first_name	id	income	last_name
0	Lauren	1592	23951.49	Murphy
1	Rebecca	27495	31019.37	Walls
2	Alejandra	19776	19058.09	Garcia
				***
12662	Mark	58060	50696.11	Torres
12663	Peter	13881	0.00	Gibson
12664	Michele	35147	19864.48	Robinson

What happens if you run the first cell one time? Two times?

	first_name	id	income	last_name
0	Lauren	1592	23951.49	Murphy
1	Rebecca	27495	31019.37	Walls
2	Alejandra	19776	19058.09	Garcia
12662	Mark	58060	50696.11	Torres
12663	Peter	13881	0.00	Gibson
12664	Michele	35147	19864.48	Robinson



What happens if you run the first cell one time? Two times?

first_name	id	income	last_name
Lauren	1592	23951.49	Murphy
Rebecca	27495	31019.37	Walls
Alejandra	19776	19058.09	Garcia
Mark	58060	50696.11	Torres
Peter	13881	0.00	Gibson
Michele	35147	19864.48	Robinson
	Lauren Rebecca Alejandra Mark Peter	Lauren 1592 Rebecca 27495 Alejandra 19776 Mark 58060 Peter 13881	Lauren       1592       23951.49         Rebecca       27495       31019.37         Alejandra       19776       19058.09              Mark       58060       50696.11         Peter       13881       0.00

```
df_income = df_income.drop(['first_name'], axis=1)

df_income = df_income.drop(['last_name'], axis=1)

[ ]: df_income
```

-	first_name	id	income	last_name	<pre>df_income = df_income</pre>	<pre>.drop(['first_name'], axis=1)</pre>
0	Lauren	1592	23951.49	Murphy		
1	Rebecca	27495	31019.37	Walls	Edited to ->	<pre>df_income = df_income.drop(['last_name'],</pre>
2	Alejandra	19776	19058.09	Garcia		
					In []: df income	
12662	Mark	58060	50696.11	Torres	<del>-</del>	
12663	Peter	13881	0.00	Gibson		
12664	Michele	35147	19864.48	Robinson		

Be very careful when editing cells that mutate variables.

## Okay, so I how do not screw things up?

- Avoid mutation until absolutely necessary!
  - Use temporary variables to work around this.
- If a cell has code that results in mutation, only run it once.
  - If you need to run it again (e.g. because of a bug), run all cells above it first.
- Restart kernel and run all cells often, and especially before you turn in your assignment.

## What's the deal with brackets?

- Why do I need brackets? When do I use parentheses and when do I use brackets?
- Why do I sometimes put strings in brackets but other times an expression?
- Why do I sometimes need double brackets??

## Use brackets when taking slices (subsets) of a DF

Key idea: Only one value goes into the brackets.

	Candidate	Party	%	Year	Result
0	Obama	Democratic	52.9	2008	win
1	McCain	Republican	45.7	2008	loss
2	Obama	Democratic	51.1	2012	win
3	Romney	Republican	47.2	2012	loss
4	Clinton	Democratic	48.2	2016	loss
5	Trump	Republican	46.1	2016	win

### How do I grab a single column?

#### elections["Candidate"].head(6) Reagan Carter Anderson Reagan Mondale Bush Name: Candidate, dtype: object

This is a Series!

### How do I grab multiple columns?

ele	ections[[ˈ	'Candidate",	"Party"]].head(6)
W-7	Candidate	Party	
0	Reagan	Republican	
1	Carter	Democratic	
2	Anderson	Independent	
3	Reagan	Republican	
4	Mondale	Democratic	
5	Bush	Republican	This is a DF!

# Use brackets when taking slices (subsets) of a DF

	Candidate	Party	%	Year	Result
0	Obama	Democratic	52.9	2008	win
1	McCain	Republican	45.7	2008	loss
2	Obama	Democratic	51.1	2012	win
3	Romney	Republican	47.2	2012	loss
4	Clinton	Democratic	48.2	2016	loss
5	Trumn	Republican	46.1	2016	win

#### How do I grab rows?

#### elections[0:3]

	Candidate	Party	%	Year	Result
0	Reagan	Republican	50.7	1980	win
1	Carter	Democratic	41.0	1980	loss
2	Anderson	Independent	6.6	1980	loss

elections[elections['Party'] == 'Independent']

	Candidate	Party	%	Year	Result
2	Anderson	Independent	6.6	1980	loss
9	Perot	Independent	18.9	1992	loss
12	Perot	Independent	8.4	1996	loss

This is a DF!

Whoa, what's going on here?

## Demo with Elections Data

Demo notebook (Elections.ipynb) and data are available on datahub

## Your Turn: Work on

D4 casestudy\_eda