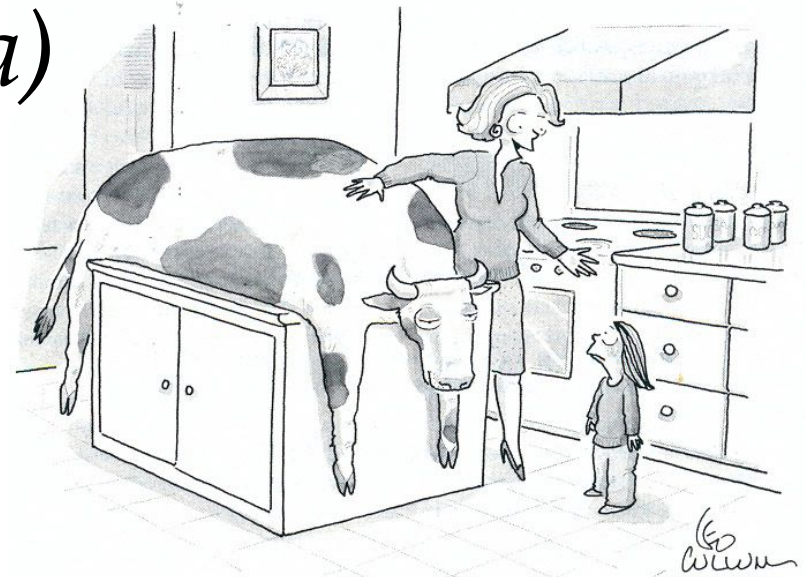




Exploring a Database (and exposing data)

*You'll need access to
Jupyter today*



"Mommy wants you to know where your food comes from."



Revisit NFL.db from last lecture

Jupyter Play Last Checkpoint: 09/12/2019 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [1]: `import iSQL`
`R = iSQL.parser("NFL.db")`

SQL: `.schema`

Submit prev next

```
CREATE TABLE Draft(  
  pid INTEGER PRIMARY KEY,  
  year INTEGER,  
  round INTEGER,  
  overall INTEGER,  
  tid INTEGER,  
  FOREIGN KEY(tid) REFERENCES Team(tid)  
)  
CREATE TABLE Game(  
  season INTEGER,  
  week TEXT,  
  date DATE,  
  vtid INTEGER,  
  vscore INTEGER,  
  htid INTEGER,  
  hscore INTEGER,  
  notes TEXT,  
  FOREIGN KEY(vtid) REFERENCES Team(tid),  
  FOREIGN KEY(htid) REFERENCES Team(tid),  
  UNIQUE(season,week,htid)  
)  
CREATE TABLE PlayedFor(  
  pid INTEGER,  
  tid INTEGER,
```

You should probably install a local or Azure version of Jupyter if you are still having problems with our cloud-based Jupyter Hub.

See course website for instructions.



Simple Questions

- ❖ How many Players?

```
SELECT COUNT(*)  
FROM Player
```

- ❖ How many Teams?

```
SELECT COUNT(*)  
FROM Team
```

- ❖ How many games?

```
SELECT COUNT(*)  
FROM Team
```

- ❖ How many records are there of a Player playing for a team in a particular season.

```
SELECT COUNT(*)  
FROM PlayedFor
```



A more advanced query

- ❖ List the names, roster year, and games played by all Carolina Panther Quarterbacks who started a game.

```
SELECT P.name, R.year, R.games, R.starts
FROM Player P, PlayedFor R, Team T
WHERE P.pid=R.pid AND R.tid=T.tid
      AND T.mascot='panthers' AND R.position = 'QB'
      AND R.starts > 0 AND R.starts <> '--
```

- ❖ There are several repeated names here. How would you get a list of DISTINCT Panther quarterbacks who've started a game?
- ❖ Same question but, in addition to their name the first year that they started?



Another query

- ❖ Give the year and locations of teams calling themselves "Colts"

```
SELECT  L.year, L.place, T.mascot
FROM    Team T, TeamLocation L
WHERE   T.tid=L.tid AND T.mascot='colts'
ORDER  BY L.year
```

- ❖ Give the year and mascots of teams based in "Baltimore". Then try "Los Angeles".
- ❖ How many different teams have called Los Angeles home?



Queries Continued

- ❖ List in order the picks from the 2017 NFL draft. Include the player's name, their college, draft round, draft order, and the team that drafted them.

```
SELECT P.name, P.college, D.round, D.overall, T.mascot
FROM Player P, Draft D, Team T
WHERE P.pid=D.pid AND T.tid=D.tid AND D.year=2017
ORDER BY D.overall
```

- ❖ 229 vs 253. What is the highest picked missing (lowest overall) draftee? Why are they missing?



Who was #10?

Rnd.	Pick #	NFL team	Player	Pos.	College	Conf.	Notes
1	1	Cleveland Browns	Myles Garrett †	DE	Texas A&M	SEC	
1	2	Chicago Bears	Mitchell Trubisky †	QB	North Carolina	ACC	from San Francisco [R1 - 1]
1	3	San Francisco 49ers	Solomon Thomas	DE	Stanford	Pac-12	from Chicago [R1 - 2]
1	4	Jacksonville Jaguars	Leonard Fournette	RB	LSU	SEC	
1	5	Tennessee Titans	Corey Davis	WR	Western Michigan	MAC	from LA Rams [R1 - 3]
1	6	New York Jets	Jamal Adams †	S	LSU	SEC	
1	7	Los Angeles Chargers	Mike Williams	WR	Clemson	ACC	
1	8	Carolina Panthers	Christian McCaffrey	RB	Stanford	Pac-12	
1	9	Cincinnati Bengals	John Ross	WR	Washington	Pac-12	
1	10	Kansas City Chiefs	Patrick Mahomes †	QB	Texas Tech	Big 12	from Buffalo [R1 - 4]
1	11	New Orleans Saints	Marshon Lattimore †	CB	Ohio State	Big Ten	
1	12	Houston Texans	Deshaun Watson †	QB	Clemson	ACC	from Philadelphia via Cleveland [R1 - 5]
1	13	Arizona Cardinals	Haason Reddick	LB	Temple	The American	
1	14	Philadelphia Eagles	Derek Barnett	DE	Tennessee	SEC	from Minnesota [R1 - 6]
1	15	Indianapolis Colts	Malik Hooker	S	Ohio State	Big Ten	
1	16	Baltimore Ravens	Marlon Humphrey	CB	Alabama	SEC	
1	17	Washington Redskins	Jonathan Allen	DE	Alabama	SEC	
1	18	Tennessee Titans	Adoree' Jackson	CB	USC	Pac-12	

❖ Seems like a major omission... lets fix it



Cleaning our data

- ❖ He's there

```
In [1]: import iSQL
R = iSQL.parser("NFL.db")
```

SQL: `SELECT * FROM Player
WHERE name like "% Mahomes"`

Submit prev next

	pid	name	height	weight	college	dob
0	29800	Patrick Mahomes	6-3	230	Texas Tech	1995-9-17

- ❖ But the Draft year is wrong. Try...

```
SELECT * FROM Draft  
WHERE pid=29800
```

```
SELECT * FROM Draft  
WHERE overall=10 AND year=2018
```




The fix

- ❖ Here is how we fix it

```
UPDATE Draft SET year=2017 WHERE pid=29800
```

- ❖ Now retry

```
SELECT P.name, P.college, D.round, D.overall, T.mascot  
FROM Player P, Draft D, Team T  
WHERE P.pid=D.pid AND T.tid=D.tid AND D.year=2017  
ORDER BY D.overall
```

- ❖ Next Rerun your iSQL cell, and then repeat the query. What happened?



Time to COMMIT

- ❖ No change is permanently made to a data base until it is committed!

- ❖ Redo this:

```
UPDATE Draft SET year=2017 WHERE pid=29800
```

- ❖ Now commit

```
COMMIT
```

- ❖ Now if we restart iSQL the fix will remain when we:

```
SELECT P.name, P.college, D.round, D.overall, T.mascot  
FROM Player P, Draft D, Team T  
WHERE P.pid=D.pid AND T.tid=D.tid AND D.year=2017  
ORDER BY D.overall
```

- ❖ Lack of COMMITs is a major source of bugs



More cleaning

❖ What's going on with "Mel Farr"

```

SELECT *
FROM Player P, PlayedFor R, Team T
WHERE P.pid=R.pid and R.tid = T.tid
      AND P.name = 'Mel Farr'

```

	pid	name	height	weight	college	dob	pid	tid	year	position	jersey	games	starts	tid	mascot
0	12209	Mel Farr	6-0	222	UCLA	1966-08-12	12209	1020	1967	RB	--	--	--	1020	lions
1	12209	Mel Farr	6-0	222	UCLA	1966-08-12	12209	1020	1968	RB	--	--	--	1020	lions
2	12209	Mel Farr	6-0	222	UCLA	1966-08-12	12209	1020	1969	RB	--	--	--	1020	lions
3	12209	Mel Farr	6-0	222	UCLA	1966-08-12	12209	1020	1970	RB	--	12	--	1020	lions
4	12209	Mel Farr	6-0	222	UCLA	1966-08-12	12209	1020	1971	RB	--	9	--	1020	lions
5	12209	Mel Farr	6-0	222	UCLA	1966-08-12	12209	1020	1972	RB	--	10	--	1020	lions
6	12209	Mel Farr	6-0	222	UCLA	1966-08-12	12209	1020	1973	RB	--	11	--	1020	lions
7	12209	Mel Farr	6-0	222	UCLA	1966-08-12	12209	1025	1989	RB	34	1	0	1025	rams



Let's not forget Jr.

Mel Farr

From Wikipedia, the free encyclopedia

For his son, see [Mel Farr Jr.](#)

Melvin Farr (November 3, 1944 – August 3, 2015) was an [American football](#) player and businessman.

A native of [Beaumont, Texas](#), Farr played [college football](#) as a [halfback](#) on the 1965 and 1966 [UCLA Bruins football teams](#) that were ranked No. 4 and No. 5 respectively in the final [AP Polls](#). He was selected as a consensus first-team All-American in 1966, gained over 1,000 yards from scrimmage in both 1965 and 1966, and was inducted into the [UCLA Athletics Hall of Fame](#) in 1988.

Farr was drafted by the [Detroit Lions](#) in the first round, seventh overall pick, of the 1967 [NFL Draft](#) and played seven years as a [running back](#) for the Lions. He led the Lions in both rushing and receiving in 1967, totaling 1,177 yards from scrimmage as a rookie, and was selected as the [NFL Rookie of the Year](#). He was twice selected to play in the [Pro Bowl](#), in 1967 and 1970. In a career shortened by injury, Farr gained a total of 4,446 yards from scrimmage and scored 36 touchdowns during his seven years in the NFL.^[1]

After retiring from football, Farr acquired a [Ford Motor Company](#) dealership in 1975, eventually expanding his business to 11 dealerships in five states. By 1998, Farr's automotive group was cited as the largest African-American owned company in the country. His

Mel Farr



Farr from 1965 UCLA yearbook

No. 24
Position: Running back
Personal information
Born: November 3, 1944 <div>Beaumont, Texas</div>
Died: August 3, 2015 (aged 70) <div>Detroit, Michigan</div>
Height: 6 ft 2 in (1.88 m)
Weight: 210 lb (95 kg)
Career information
High school: Hebert (Beaumont, TX)
College: UCLA
NFL Draft: 1967 / Round: 1 / Pick: 7
Career history
<ul style="list-style-type: none"> Detroit Lions (1967–1973)
Career highlights and awards

Mel Farr Jr.

From Wikipedia, the free encyclopedia

Melvin Farr Jr. (born August 12, 1966) is a former [American football](#) [running back](#) who played one season for the [Los Angeles Rams](#) of the [National Football League](#) (NFL). He attended and played [college football](#) for the [UCLA Bruins](#).

Farr is part of a family full of professional football players. Farr is the son of former NFL player [Mel Farr](#), the nephew of former [American Football League](#) and NFL player [Miller Farr](#), and the older brother of former NFL player [Mike Farr](#).

Farr is a member of [Alpha Phi Alpha](#) fraternity.^[1]

References [[edit](#)]

- ↑ ""Archived copy"". Archived from the original on 2010-04-19. Retrieved 2008-01-30.

Mel Farr Jr.

No. 34
Position: Running back
Personal information
Born: August 12, 1966 (age 53) <div>Santa Monica, California</div>
Career information
College: UCLA
NFL Draft: 1988 / Round: 9 / Pick: 248
Career history
<ul style="list-style-type: none"> Los Angeles Rams (1989)
Career NFL statistics
Games played: 1
Player stats at NFL.com [ⓘ]

How do we fix this one?



Adding a new Farr

- ❖ The player data seems to match Jr. We need add new player record for his dad.
- ❖ First we need a pid

```
SELECT MAX(pid) FROM Player
```

- ❖ Then add a new Player

```
INSERT INTO Player  
VALUES(31861, 'Mel Farr', '6-2', 210, 'UCLA', '1944-11-03')
```

- ❖ Now update the rosters. How does this work?

```
UPDATE PlayedFor SET pid=31861 WHERE pid=12209 AND tid=1020
```

- ❖ Don't forget to commit!



Errors, errors, everywhere

- ❖ Dirty data is everywhere
 - What about Tim Anderson, Marvin Powell, and Jeff Smith to name a few
 - It takes a it of sleuthing to figure them out
- ❖ However, we need to be able to write correct queries in the presence of dirty data
- ❖ This is the challenge and learning objective of your next problem set
 - I have fixed 1000s of errors that appear in the version of NFL.db that you have
 - I have probably introduced dozens more
- ❖ Your HW queries need to work correctly on my instance as well as yours.



Next Time

- ❖ Embedding SQL queries within a traditional programming language

