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

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"

```sql
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.

```sql
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?

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

❖ Seems like a major omission… let’s fix it
Cleaning our data

- He's there

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

- Now retry
  
  ```sql
  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:
  ```sql
  UPDATE Draft SET year=2017 WHERE pid=29800
  ```
- Now commit
  ```
  COMMIT
  ```
- Now if we restart iSQL the fix will remain when we:
  ```sql
  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'
```

<table>
<thead>
<tr>
<th>pid</th>
<th>name</th>
<th>height</th>
<th>weight</th>
<th>college</th>
<th>dob</th>
<th>pid</th>
<th>tid</th>
<th>year</th>
<th>position</th>
<th>jersey</th>
<th>games</th>
<th>starts</th>
<th>tid</th>
<th>mascot</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Mel Farr</td>
<td>6-0</td>
<td>222</td>
<td>UCLA</td>
<td>1966-08-12</td>
<td>12209</td>
<td>1020</td>
<td>1967</td>
<td>RB</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1020</td>
<td>lions</td>
</tr>
<tr>
<td>1</td>
<td>Mel Farr</td>
<td>6-0</td>
<td>222</td>
<td>UCLA</td>
<td>1966-08-12</td>
<td>12209</td>
<td>1020</td>
<td>1968</td>
<td>RB</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1020</td>
<td>lions</td>
</tr>
<tr>
<td>2</td>
<td>Mel Farr</td>
<td>6-0</td>
<td>222</td>
<td>UCLA</td>
<td>1966-08-12</td>
<td>12209</td>
<td>1020</td>
<td>1969</td>
<td>RB</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>1020</td>
<td>lions</td>
</tr>
<tr>
<td>3</td>
<td>Mel Farr</td>
<td>6-0</td>
<td>222</td>
<td>UCLA</td>
<td>1966-08-12</td>
<td>12209</td>
<td>1020</td>
<td>1970</td>
<td>RB</td>
<td>--</td>
<td>12</td>
<td>--</td>
<td>1020</td>
<td>lions</td>
</tr>
<tr>
<td>4</td>
<td>Mel Farr</td>
<td>6-0</td>
<td>222</td>
<td>UCLA</td>
<td>1966-08-12</td>
<td>12209</td>
<td>1020</td>
<td>1971</td>
<td>RB</td>
<td>--</td>
<td>9</td>
<td>--</td>
<td>1020</td>
<td>lions</td>
</tr>
<tr>
<td>5</td>
<td>Mel Farr</td>
<td>6-0</td>
<td>222</td>
<td>UCLA</td>
<td>1966-08-12</td>
<td>12209</td>
<td>1020</td>
<td>1972</td>
<td>RB</td>
<td>--</td>
<td>10</td>
<td>--</td>
<td>1020</td>
<td>lions</td>
</tr>
<tr>
<td>6</td>
<td>Mel Farr</td>
<td>6-0</td>
<td>222</td>
<td>UCLA</td>
<td>1966-08-12</td>
<td>12209</td>
<td>1020</td>
<td>1973</td>
<td>RB</td>
<td>--</td>
<td>11</td>
<td>--</td>
<td>1020</td>
<td>lions</td>
</tr>
<tr>
<td>7</td>
<td>Mel Farr</td>
<td>6-0</td>
<td>222</td>
<td>UCLA</td>
<td>1966-08-12</td>
<td>12209</td>
<td>1025</td>
<td>1989</td>
<td>RB</td>
<td>34</td>
<td>1</td>
<td>0</td>
<td>1025</td>
<td>rams</td>
</tr>
</tbody>
</table>
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 personal information:

- Born: November 3, 1944
- Died: August 3, 2015 (aged 70)
- Detroit, Michigan
- Height: 6 ft 2 in (1.88 m)
- Weight: 210 lb (95 kg)

Career information:

- NFL Draft: 1967 / Round: 1 / Pick: 7
- Career highlights and awards:
  - Detroit Lions (1967–1973)

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 ]


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
  
  ```sql
  SELECT MAX(pid) FROM Player
  ```
- Then add a new Player
  
  ```sql
  INSERT INTO Player
  VALUES(31861, 'Mel Farr', '6-2',210,'UCLA','1944-11-03')
  ```
- Now update the rosters. How does this work?
  
  ```sql
  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