Normalizing Our Data

First Normal Form

- Fully atomized

PersonNames	Dogs	Dogs2	Breeds
Bubba Humbucker	Chief, Larry, Butch, Edna	Two-Ton, Poodlemeister	Mutt, Retriever, Black Lab, German Shepherd, Poodle

FirstName	LastName	Dog	Breed
Bubba	Humbucker	Chief	Mutt
Bubba	Humbucker	Larry	Retriever
Bubba	Humbucker	Butch	Black Lab
Bubba	Humbucker	Edna	German Shepherd
Bubba	Humbucker	Two-Ton	Black Lab
Bubba	Humbucker	Poodlemeister	Poodle

First Normal Form

- No repeating groups

FirstName	LastName	Dog	Breed
Bubba	Humbucker	Chief	Mutt
Bubba	Humbucker	Larry	Retriever
Bubba	Humbucker	Butch	Black Lab
Bubba	Humbucker	Edna	German Shepherd
Bubba	Humbucker	Two-Ton	Black Lab
Bubba	Humbucker	Poodlemeister	Poodle

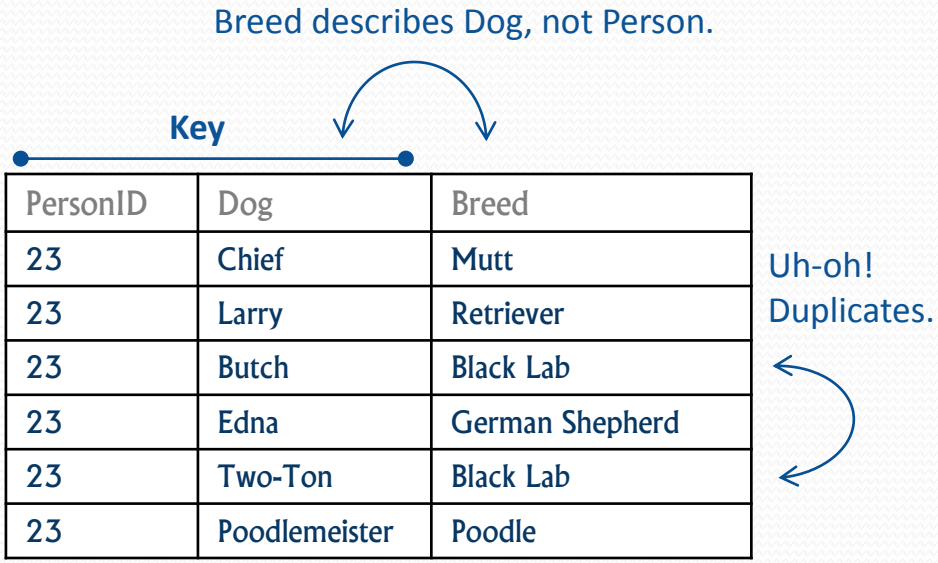
PersonID	FirstName	LastName
23	**Bubba**	**Humbucker**

Key

PersonID	Dog	Breed
23	Chief	Mutt
23	Larry	Retriever
23	Butch	Black Lab
23	Edna	German Shepherd
23	Two-Ton	Black Lab
23	Poodlemeister	Poodle

Second Normal Form

- Each column must depend on the **entire** primary key (multi-field keys.)



Second Normal Form

- Each column must depend on the **entire** primary key (esp. multi-field keys.)

Key

PersonID	DogID
23	321
23	345
23	567
23	639
23	782
23	784

Key

DogID	DogName	BreedID
321	Chief	1
345	Larry	2
567	Butch	3
639	Edna	4
782	Two-Ton	3
784	Poodlemeister	6

Key

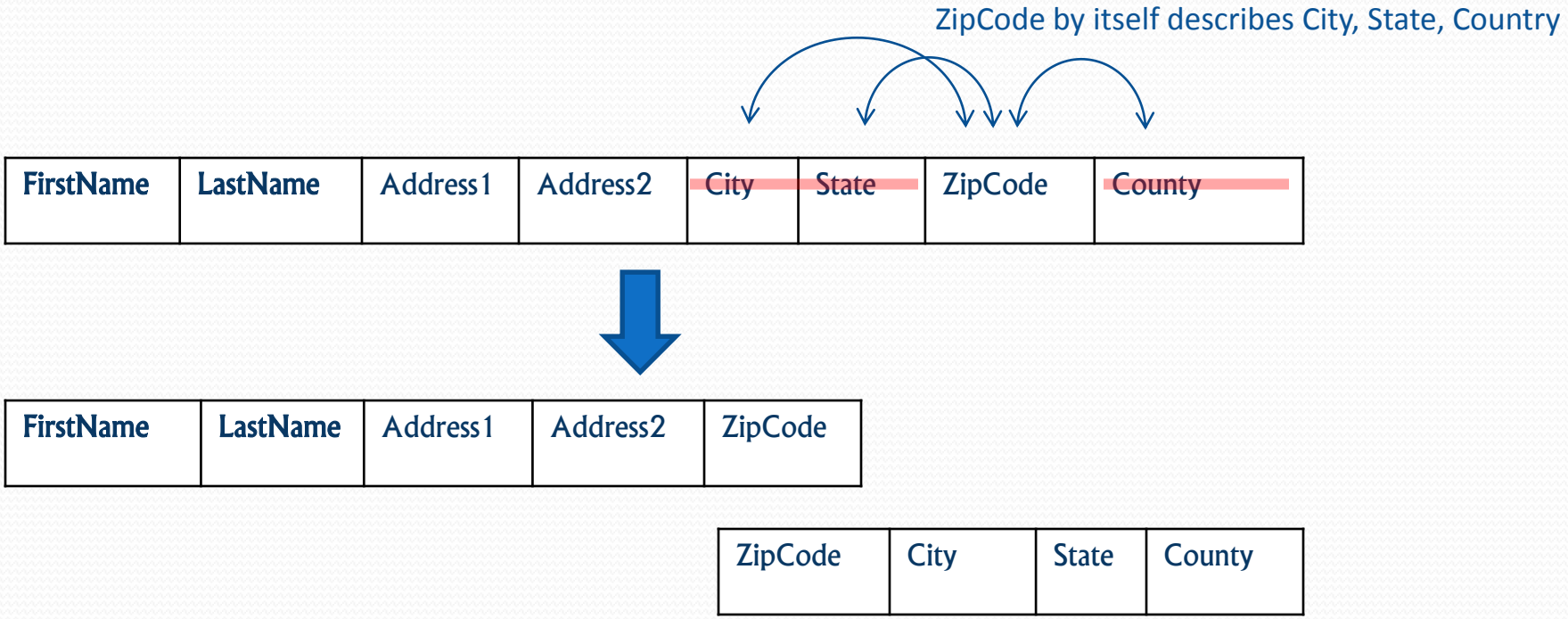
BreedID	Breed
1	Mutt
2	Retriever
3	Black Lab
4	German Shepherd
5	Poodle

Third Normal Form

- Each column must depend on **directly** on the primary key. (NF1)
- Attributes that do not contribute to the description of the primary key are removed from the table. (NF2)
- And…

Third Normal Form

- No transitive functional dependency



"Every attribute must provide a fact about the key (NF1), the whole key (NF2), and nothing but the key(NF3)….so help me Codd."

Remove Further Redundancy

- Remove attributes that can be calculated/derived from other ones
  - Averages, sums, percentages, age…

Age = NOW() - BirthDate / 365

The computer knows “NOW” and that “365” is the number of days in a year, so all we need to store in the DB is the “BirthDate” in order to calculate someone’s age.

Relational Integrity

- Using the database design to insure links between tables
- Requires a value to exist in parent or child
- Avoids child records being orphaned
- UPDATES & DELETES
  - Restrict?
  - Cascade?