

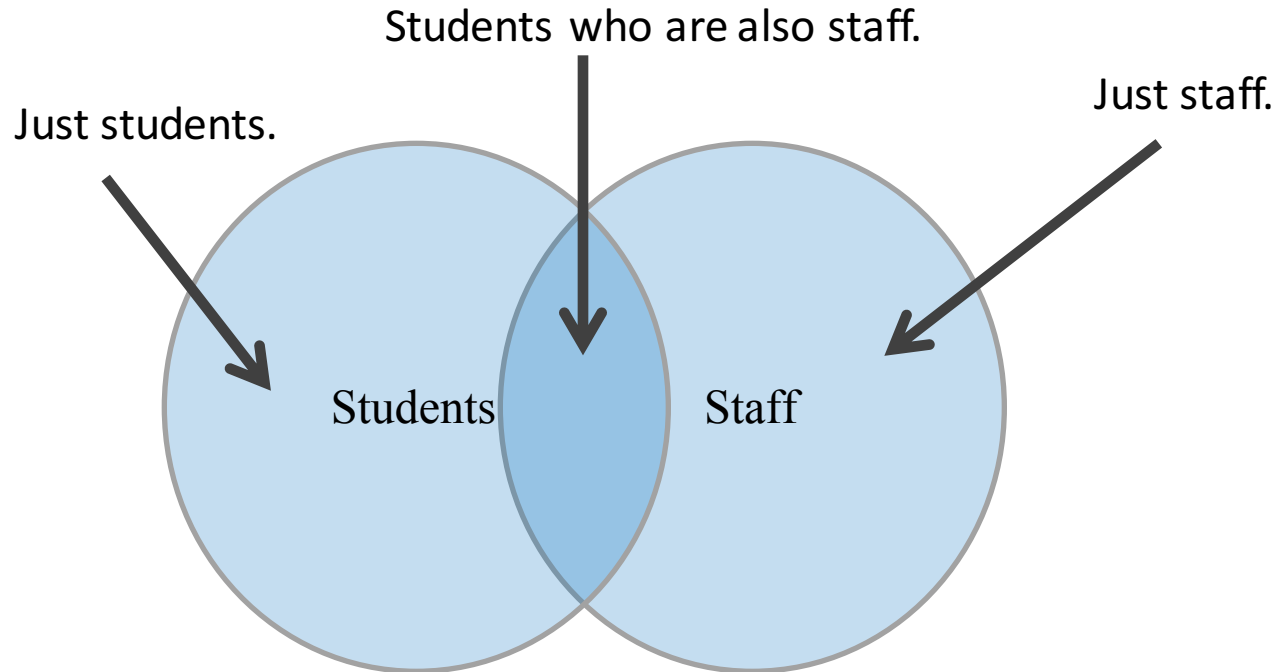
Pandas Data Structures

- Series Object (1 dimensional, a row)
- DataFrame Object (2 dimensional, a table)
- Querying
 - *iloc[], for querying based on position*
 - *loc[], for querying rows based on label*
 - *Querying the DataFrame directly*
 - *Projecting a subset of columns*
 - *Using a boolean mask to filter data*

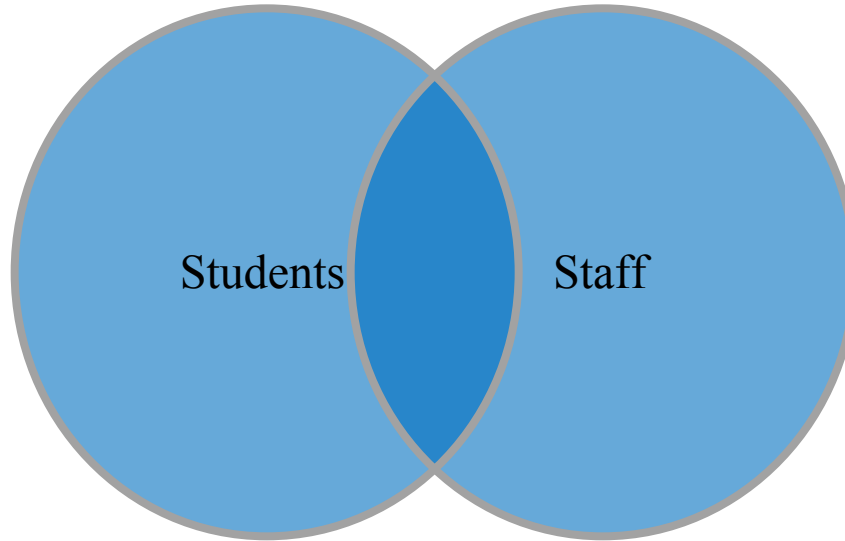
Setting Data in Pandas

- **To add new data**
 - `df[column]=[a,b,c]`
- **To set default data (or overwrite all data):**
 - `df[column]=2`

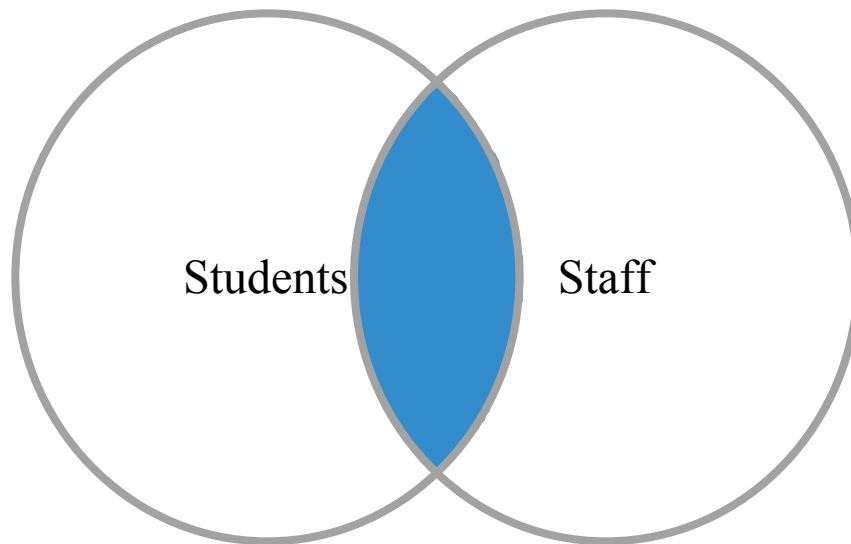
Venn Diagram



Full outer join (union)



Inner join (intersection)



- **Chain Indexing:**
 - `df.loc[“Washtenaw”][“Total Population”]`
 - *Generally bad, pandas could return a copy of a view depending upon numpy*
- **Code smell**
 - *If you see a `][` you should think carefully about what you are doing (Tom Augspurger)*

(a,b) (c,d): Scales

- **Ratio scale:**
 - *units are equally spaced*
 - *mathematical operations of +-/• are all valid*
 - *E.g. height and weight*
- **Interval scale:**
 - *units are equally spaced, but there is no true zero*
- **Ordinal scale:**
 - *the order of the units is important, but not evenly spaced.*
 - *Letter grades such as A+, A are a good example*
- **Nominal scale:**
 - *categories of data, but the categories have no order with respect to one another.*
 - *E.g. Teams of a sport.*