Python
Data
Structure
Review

→ Tables

Attributes (columns)

Items (rows)

Cell containing value
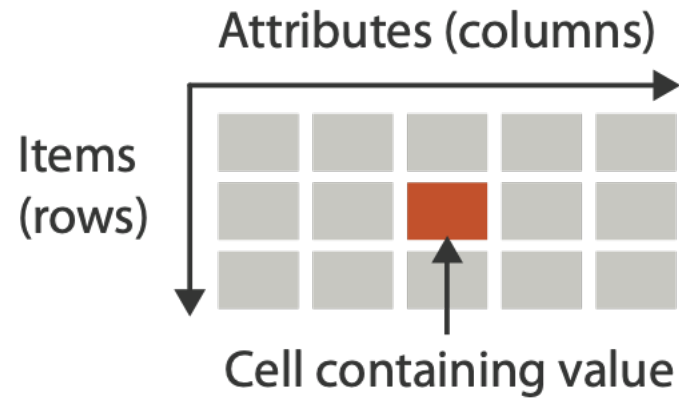
- Tables in Python are well represented using the pandas package.

- A flat table has
  - One item/record per row
  - Each column is an attribute
 - Cell holds value for item-attribute pair

Attributes / Features / Variables

Items / Objects / Records / Observations / Data Points

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

# Useful built-in Data Types

- Sometimes nothing works better than the built-in types!

  - Dictionary
  - Tuple
  - List

- Remember, python loves its iterators, so use those whenever possible

Python Dictionary Example - Studyopedia.com
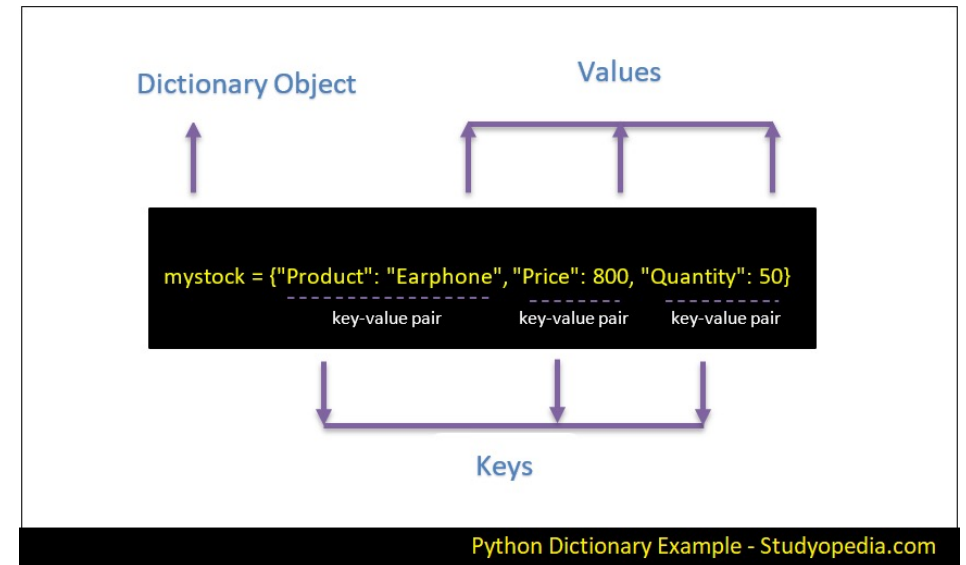
Pandas DataFrame

- 2-dimensional data structure that can store data of different types (including characters, integers, floating point values, categorical data and more) in columns. It is similar to a spreadsheet, a SQL table or the data.frame in R.

- Each column in a DataFrame is a Series

- If the data is numeric, it uses numpy under the hood and can be used for matrix math
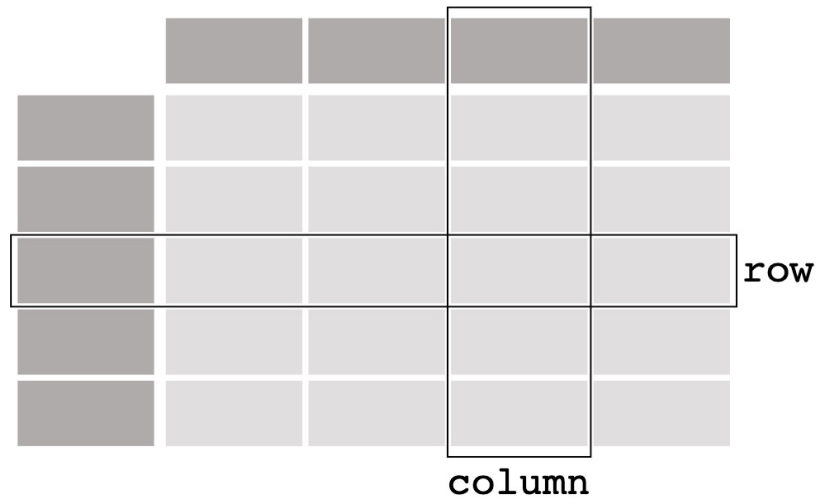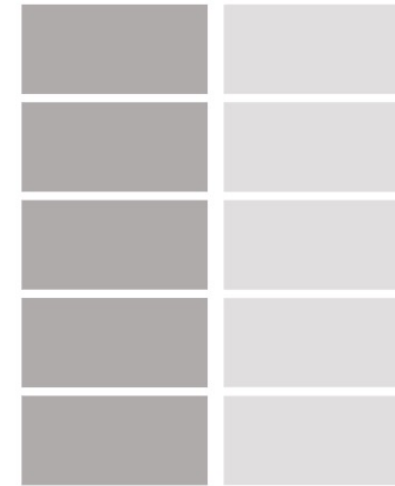
Selecting Subsets of DataFrames

- When selecting subsets of data, square brackets [] are used.
- Inside these brackets, you can use a single column/row label, a list of column/row labels, a slice of labels, a conditional expression or a colon.
- Select specific rows and/or columns using loc when using the row and column names
- Select specific rows and/or columns using iloc when using the positions in the table
- You can assign new values to a selection based on loc/iloc.

What is "Data"

- Physical type in memory
  - float64
  - uint16
  - Etc…

- Abstract type
  - Nominal
  - Ordinal
  - Interval
  - Ratio
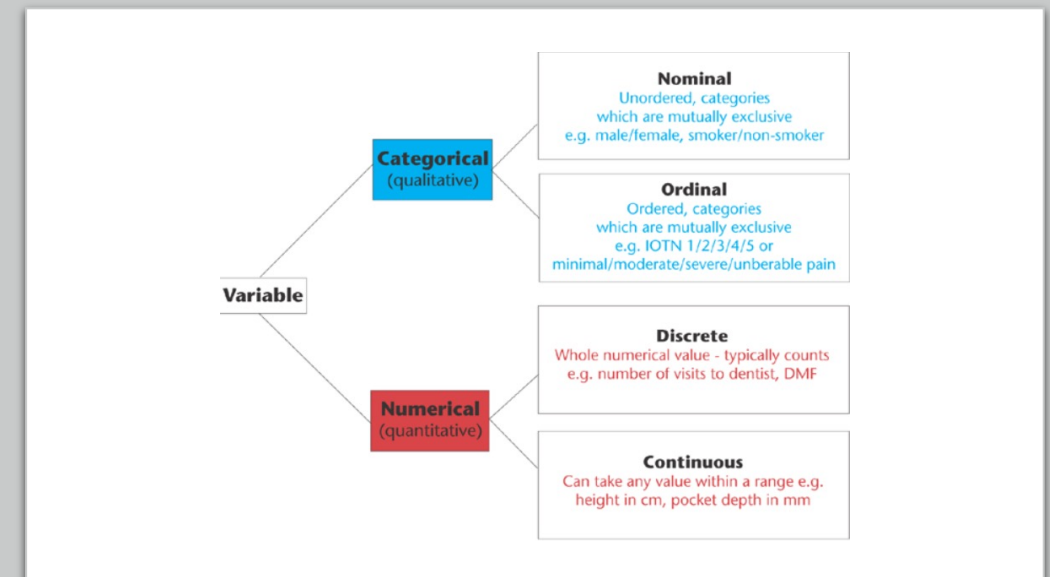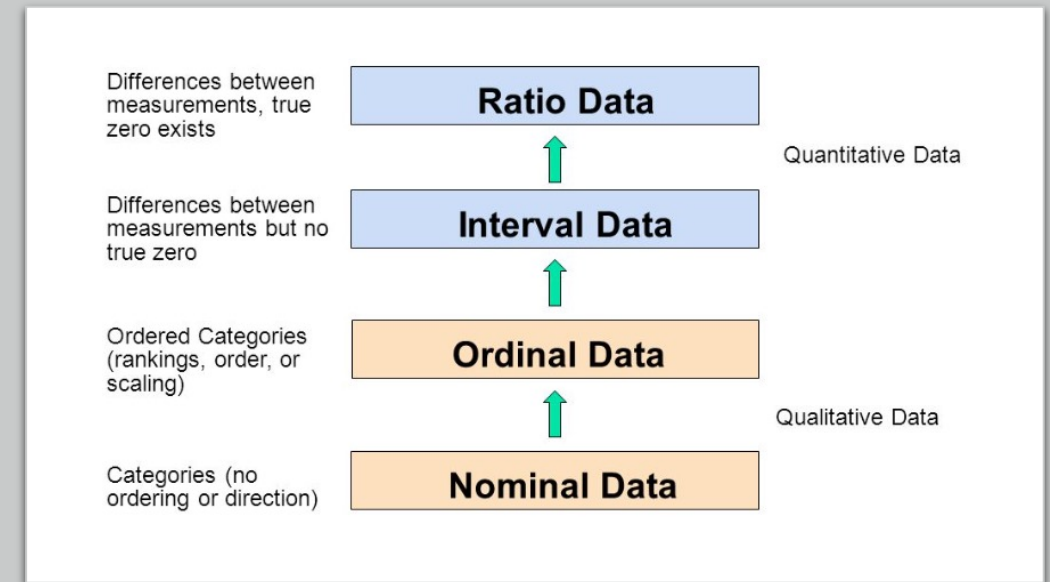
Pop Quiz

- Each scale is represented once in the list below.

  - Favorite candy bar
  - Weight of luggage
  - Year of your birth
  - Egg size (small, medium, large, extra large, jumbo)

Pop Quiz

- Each scale is represented once in the list below.

    - Military rank
    - Number of children in a family
    - Jersey numbers for a football team
    - Shoe size
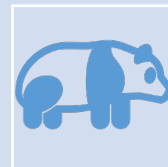
Break into pairs for HW0

Find a partner

Ask questions if you have any problems

Go to Canvas and get the Python Pandas Group Lab jupyter notebook and data file

Take the Python Pandas Quiz on canvas (OK if you discuss the quiz as a group, but please don't post the answers in a public forum)

Turn in the ipynb file as a group HW0