

BC COMS 1016: Intro to Comp Thinking & Data Science

Lecture 4 More Tables

Reminders



- HW00 due tonight
 - Individual assignment
 - Ask questions in slack #homeworks
 - Answer your peers questions as well
 - Can count for participation grade
- Lab01 due Friday night
- HW01 – Released today, due Monday night

Grades for Labs



- Lab00
 - Participation grade
- Labs01 ... 12
 - Grade will be based on tests on gradescope



- Exploration **Week 1 - 2**
 - Introduction to Python
 - Working with data
- Inference **Week 3 - 5**
 - Probability
 - Statistics
- Prediction **Week 6-7**
 - Machine Learning
 - Regression & Classification



- No lecture Monday (11/02) & Tuesday (11/03)
- Office Hours cancelled
 - I'm happy to open up office hours during the day

A black and white photograph of the exterior of a classical building, likely Barnard College. The building features large, ornate Corinthian columns supporting a series of classical cornices. The word "BARNARD" is visible in raised letters on the side of the building. The sky is clear and blue.

Tables



Table Structure

- A Table is a sequence of labeled columns
- Row: represents one individual
- Column: represents one attribute of the individuals

Name	Code	Area (m2)
California	CA	163696
Nevada	NV	110567



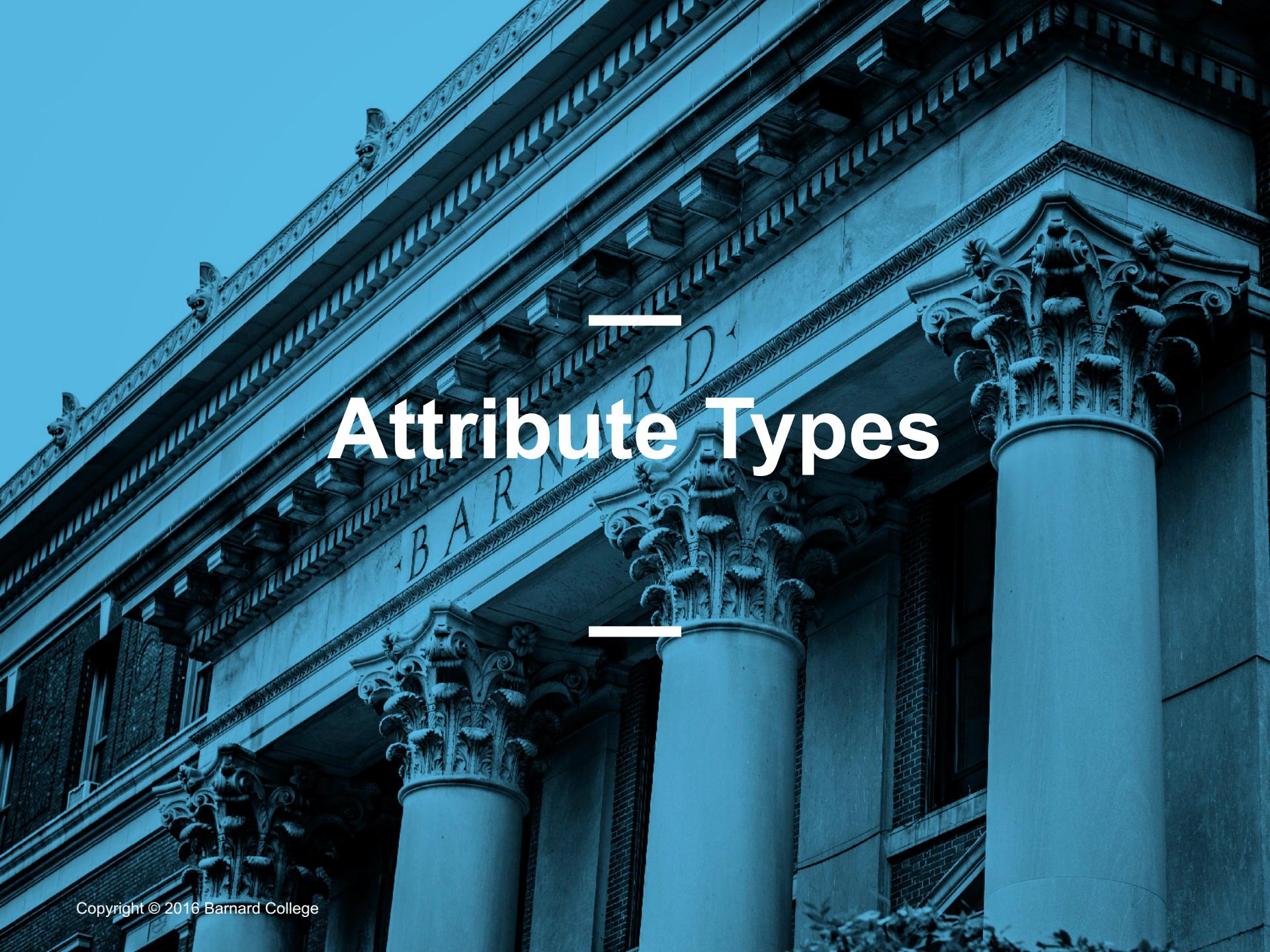
Tables – select and drop

- `.select(<Column Name>)`
 - Returns a new table with the specified columns
- `.select(<Int i>)`
 - Returns a new table with the column at index *i*
- `drop(<Column Name>)`
 - Returns a new table without the specified columns
- `.drop(<Int i>)`
 - Returns a new table without the column at index *i*

Some Table Operations



- `t.sort(label)` – constructs a new table with rows sorted by the specified column
- `t.where(label, condition)` – constructs a new table with just the rows that match the condition
- More are listed at
<http://coms1016.barnard.edu/python-reference.html>



Attribute Types



All values in a column of a table should be both the same type and be comparable to each other

- **Numerical** – values are from a numerical scale
 - Numerical measurements are ordered
 - Differences are meaningful
- **Categorical** – values from a fixed inventory
 - May or may not have an ordering
 - Categories are the same or different



Values as numbers are not guaranteed to be numerical

- Census example: SEX code (0, 1, 2)
- Arithmetic on these “numbers” is meaningless
- The variable SEX is still categorical, even though numbers were used for the categories



Census Data

The Decennial Census



- Every ten years, Census Bureau counts how many people there are in the U.S.
- Census Bureau estimates how many people are in US during the other 9 years
- U.S. Constitution Article 1, Section 2:
 - “Representatives and direct Taxes shall be apportioned among the several States ... according to their respective Numbers ...”



- <https://www2.census.gov/programs-surveys/popest/datasets/>
- <https://www2.census.gov/programs-surveys/popest/datasets/2010-2015/national/totals/>
- demo