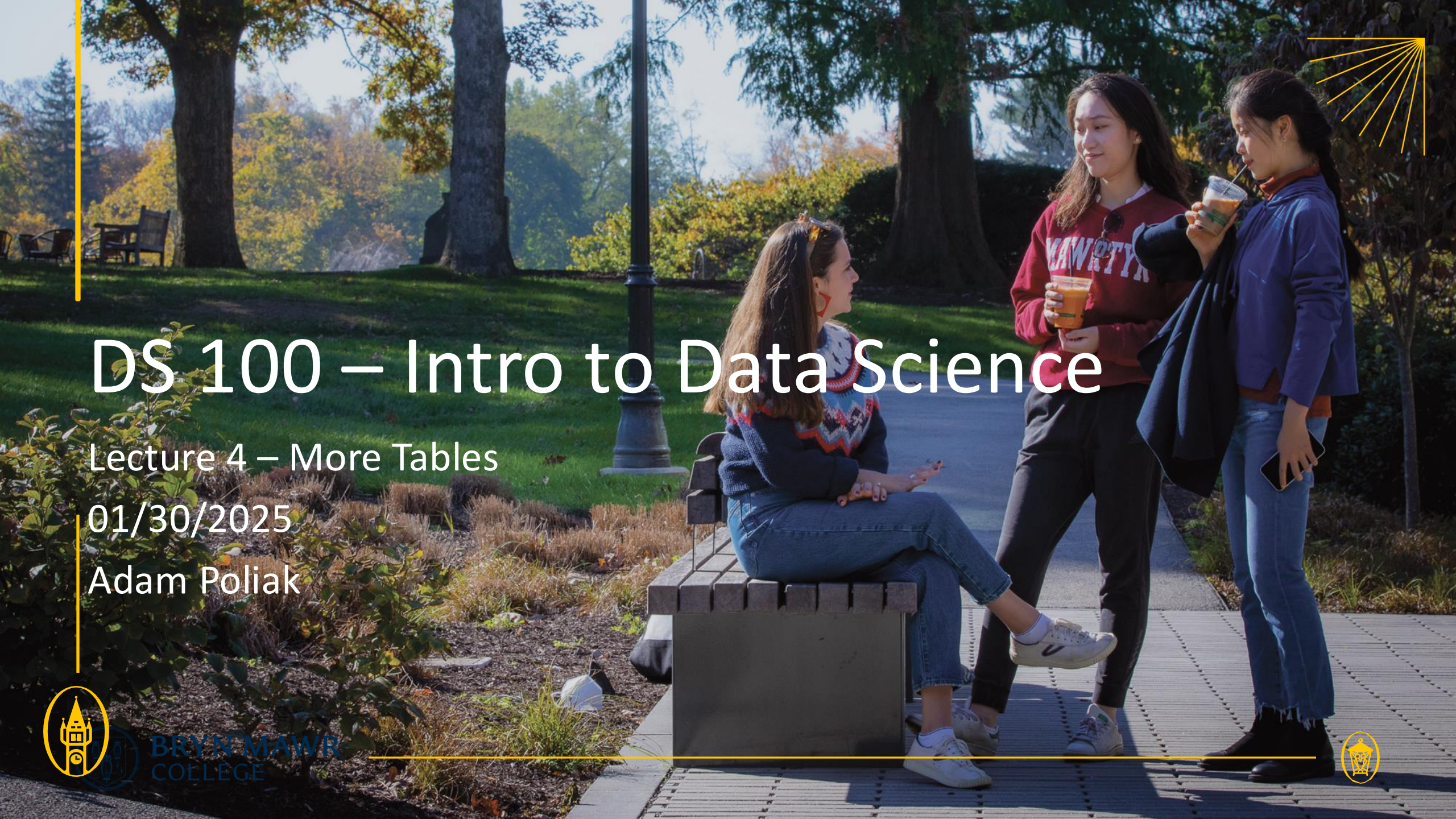

DS 100 – Intro to Data Science

Lecture 4 – More Tables

01/30/2025

Adam Poliak

BRYN MAWR COLLEGE

Announcements

HW 00

- Expect feedback over the weekend (graded already but waiting for late days)

Lab 01

- Due Friday (01/31)

HW01

- Due Wednesday (02/05)

TAs

Patrick Kelly

- Senior, Political Science Major

Allison Liao

- Junior, CS major

Candy Li

- Senior, Psych major

Office Hours

TA office hours in Park 230

Patrick Monday OH at HC (location TBD)

Allison's Tuesday time TBD

Adam's OH:
Thursday @ Dalton 300
Friday @ Park 200C

Adam	Thursday 2:30-3:30 Friday 11:30 – 1
Allison	Sunday 6-8 Tuesday TBD
Patrick	Monday 2-4 (HC) Wednesday 4-6
Candy	Wednesday 5-7 Thursday 5-7

Course Outline

Exploration **Week 1 - 5**

- Introduction to Python
- Working with data

Inference **Week 6 - 10**

- Probability
- Statistics

Prediction **Week 11 - 14**

- Machine Learning
- Regression & Classification

Review

int

float

String

built_in_function_or_method: max, min, …

Table

Tables

Table

- 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, condiction) – constructs a new table with just the rows that match the condition

- More are listed at http://bmc-ds-100.github.io/python-reference.html

Attribute Types

Types of Attributes

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

Numerical Attributes

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 ...”

Data

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