Lecture 5

DATA 8

Spring 2022

Cal

Building tables

Announcements

- HW 2 is due Thursday, 2/3
  - Submit Wednesday for a bit of extra credit
- Lecture 6 next Monday will be broadcast from Wheeler
  - You can still join via Zoom: bit.ly/d8sp22zoom
  - To attend lecture in person on Monday:
    Please fill out bit.ly/data8lec6 by 8pm Sunday
  - There will still be videos posted to data8.org/sp22
  - (Yes, we'll try to post videos more quickly.)

Weekly Goals

- Monday:
  - Python basics
  - Tables
- Wednesday:
  - Types of data
  - Arrays
- **Today**:
  - Creating new tables
  - Manipulating tables and columns

Columns

Columns are Arrays

A table of a column is an array of values (one per row).

For a table `t`, use `t.column(label)` or `t.column(index)`

For an array `s`:

- `s.item(index)` gives the value at an index (starting at 0)
- Two ways to aggregate the values in an array:
  - `np.mean(s), np.sum(s), np.max(s), np.min(s)`
  - `s.mean(), s.sum(), s.max(), s.min()`

(Demo)

Ranges

(Demo)

Creating Tables

Ways to create a table

- **`Table.read_table(filename)`** - reads a table from a file (such as a spreadsheet)
- **`Table()`** - an empty table to which columns are added
- and… **`select, where, sort, drop, take`** all create new tables based on existing tables

(Demo)

Ranges

A range is an array of consecutive numbers

- `np.arange(end)`:
  An array of increasing integers from 0 up to `end`

- `np.arange(start, end)`:
  An array of increasing integers from `start` up to `end`

- `np.arange(start, end, step)`:
 A range with `step` between consecutive values

The range always includes `start` but excludes `end`

Example

W. E. B. Du Bois, 1868-1963

- Scholar, historian, activist, and data scientist
- A founder of the NAACP
- Made a series of visualizations for the 1900 Paris Exposition
  - Goal: educate people about the lives of Black Americans
  - Hundreds of photographs and patents
  - About 30 handmade statistical graphics (created in 3 months)
  - Now in the Library of Congress
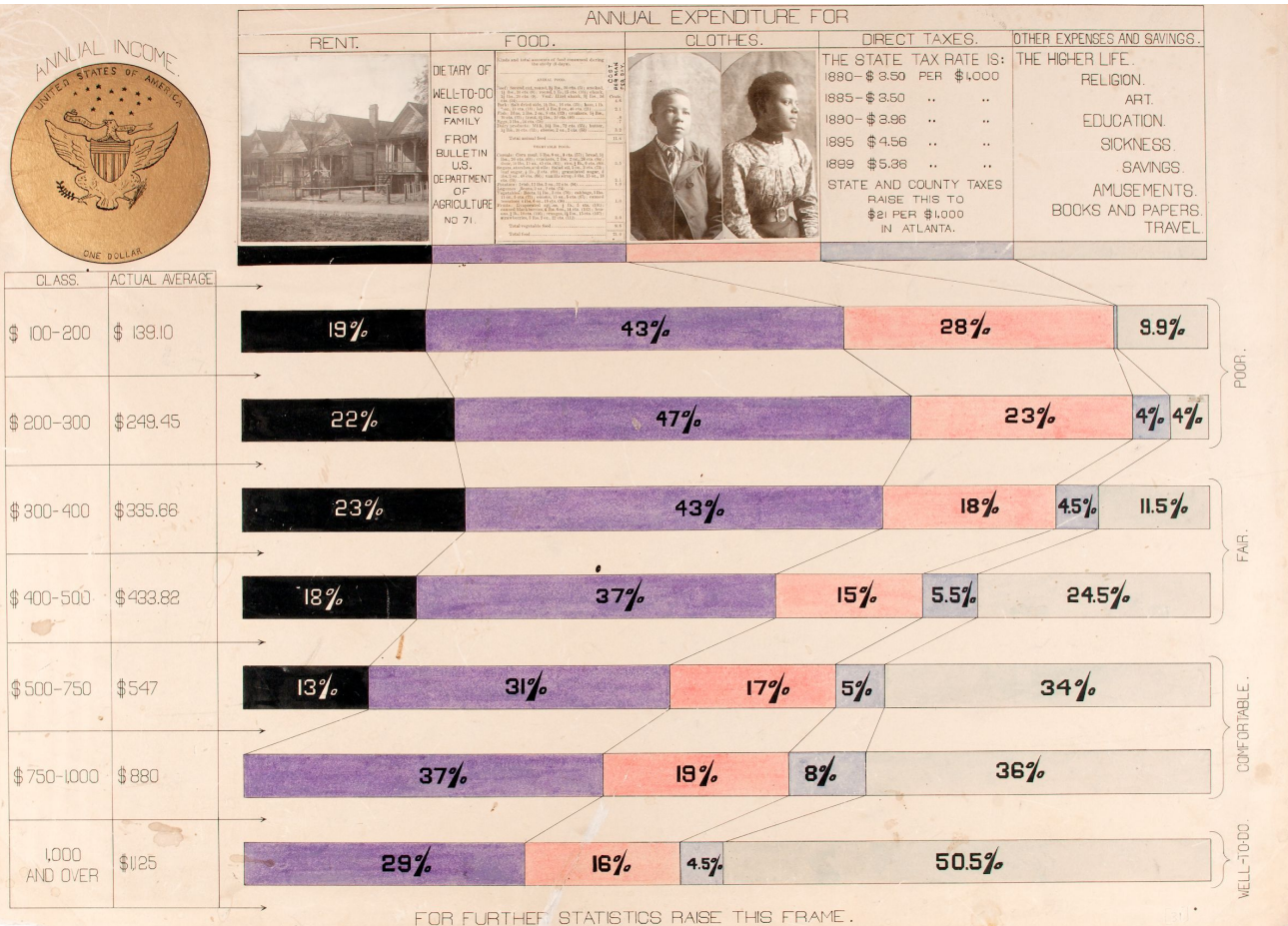
This is a photograph of the actual graph produced by Du Bois and his graduate students. The title (omitted) contains language that was used at the time, but would not be appropriate to use today.

(Demo)

Discussion Question

Use the table functions we learned this week to find the income bracket ("class") that spent the highest percentage of their income on rent.

Table Methods So Far

- Creating tables: `Table().with_columns` ; `Table.read_table`
- Finding the size: `t.num_rows` and `t.num_columns`
- Listing/changing the column labels: `t.labels` and `t.relabeled`
- An array of column values: `column` takes a label or index
- A table containing some of the original columns: `select`, `drop`
- A table containing some of the original rows: `where`, `take`
- A table with the original rows sorted: `sort`

What's next? More ways of using `where` to focus on certain rows.

Optional Material: Printing and Strings

(Demo)