YaleNUSCollege

YSC2239 Lecture 3

Today’s class

- Arrays
- Creating new tables
- Manipulating columns of tables

- Reading: Chapter 5, 6.1, 6.2

Arrays

Arrays

An array contains a sequence of values

- All elements of an array should have the same type
- Arithmetic is applied to each element individually
- Adding arrays adds elements (if same length!)
- A column of a table is an array

(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`**

Building Tables

Ways to create a table

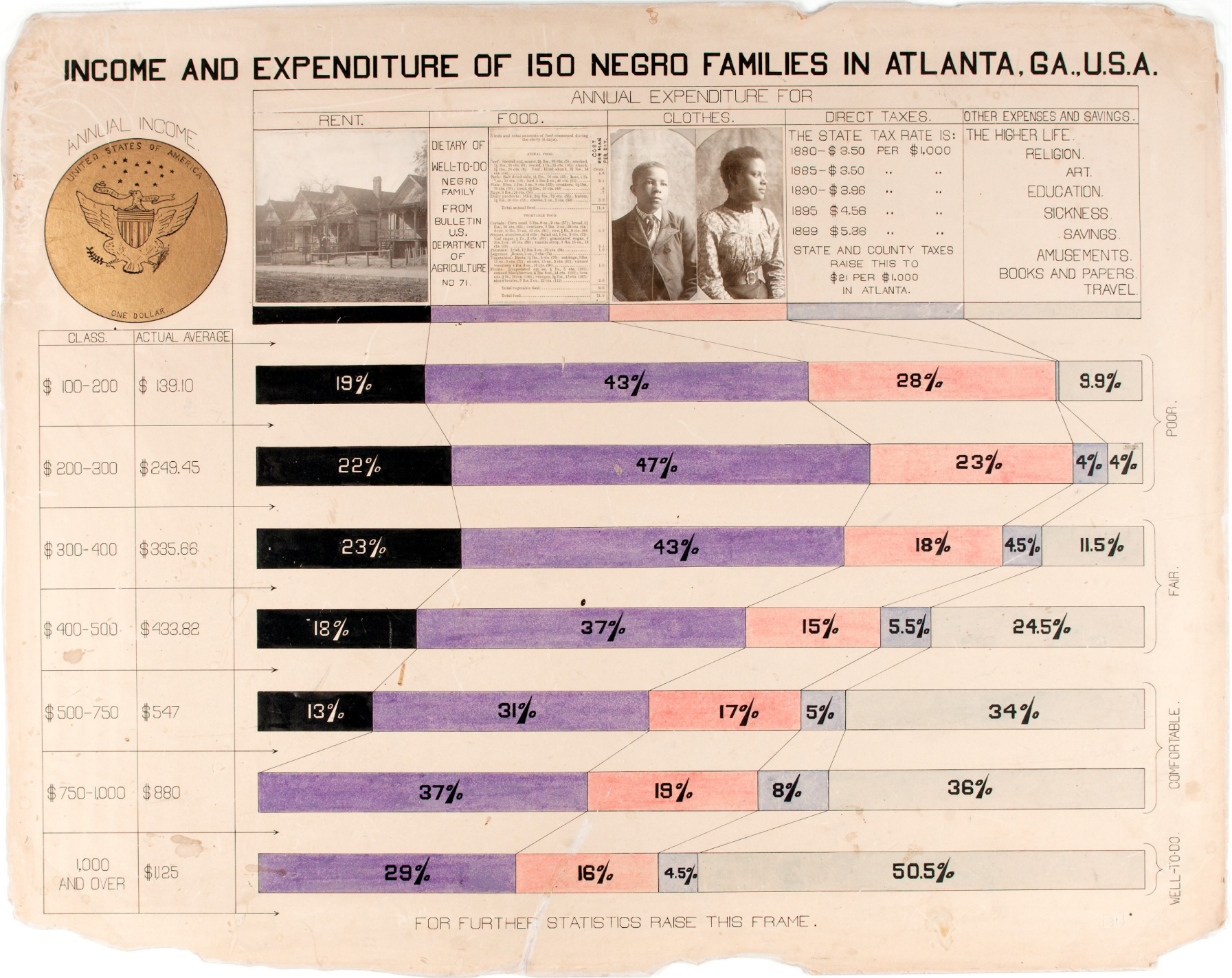
- `Table.read_table(filename)` - reads a table from a spreadsheet
- `Table()` - an empty table

- and… `select`, `where`, `sort` and so on all create new tables

Example

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

- Scholar, historian, activist, and data scientist
- NAACP founder
- Made a series of visualizations for the 1900 Paris Exposition
  - Goal: change the way people see Black Americans
  - Hundreds of photographs and patents
  - 60+ handmade graphs in 3 months


INCOME AND EXPENDITURE OF 150 NEGRO FAMILIES IN ATLANTA, GA., U.S.A.
ANNUAL INCOME.
UNITED STATES OF AMERICA
ONE DOLLAR
ANNUAL EXPENDITURE FOR
RENT.
FOOD.
CLOTHES.
DIRECT TAXES.
OTHER EXPENSES AND SAVINGS.
DIETARY OF WELL-TO-DO NEGRO FAMILY FROM BULLETIN U.S. DEPARTMENT OF AGRICULTURE NO 71.
THE STATE TAX RATE IS:
1880 – $3.50 PER $1,000
1885 – $3.50 .. ..
1890 – $3.96 .. ..
1895 $4.56 .. ..
1899 $5.36 .. ..
STATE AND COUNTY TAXES RAISE THIS TO $21 PER $1,000 IN ATLANTA.
THE HIGHER LIFE.
RELIGION.
ART.
EDUCATION.
SICKNESS.
SAVINGS.
AMUSEMENTS.
BOOKS AND PAPERS.
TRAVEL.
CLASS.
ACTUAL AVERAGE.
$ 100-200 $ 139.10
19% 43% 28% 9.9%
$ 200-300 $249.45
22% 47% 23% 4% 4%
$ 300-400 $335.66
23% 43% 18% 4.5% 11.5%
$ 400-500 $433.82
18% 37% 15% 5.5% 24.5%
$ 500-750 $547
13% 31% 17% 5% 34%
$ 750-1,000 $880
37% 19% 8% 36%
1,000 AND OVER $1125
29% 16% 4.5% 50.5%
POOR.
FAIR.
COMFORTABLE.
WELL-TO-DO.
FOR FURTHER STATISTICS RAISE THIS FRAME.

(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

- Creating and extending tables:
  - `Table().with_column` and `Table.read_table`
- Finding the size: `num_rows` and `num_columns`
- Referring to columns: labels, relabeling, and indices
  - `labels` and `relabeled`; column indices start at 0
- Accessing data in a column
  - `column` takes a label or index and returns an array
- Using array methods to work with data in columns
  - `item`, `sum`, `min`, `max`, and so on
- Creating new tables containing some of the original columns:
  - `select`, `drop`

(Demo)

Manipulating Rows

- `t.sort(column)` sorts the rows in increasing order
- `t.sort(column, descending=True)` sorts the rows in decreasing order
- `t.take(row_numbers)` keeps the numbered rows
  - Each row has an index, starting at 0
- `t.where(column, are.condition)` keeps all rows for which a column's value satisfies a condition
- `t.where(column, value)` keeps all rows for which a column's value equals some particular value
  - Same as `t.where(column, are.equal_to(value))`

Discussion Questions

The table **nba** has columns **PLAYER**, **POSITION**, and **SALARY**.

a) Create an array containing the names of all point guards (**PG**) who make more than $15M/year

```
guards = nba.where('POSITION', 'PG')
guards.where('SALARY', are.above(15)).column('PLAYER')
```

b) How to combine two tables into one?

```
nba.where('POSITION','PG') .append(nba.where('POSITI
ON','C').show(100)
```

(Demo)

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 in some way

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

To do

- Lab 2 already posted on Canvas and due on Wednesday