DATA 8

Spring 2020

Lecture 5

Building tables

Weekly Goals

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

Announcements

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`**

Ways to create a table

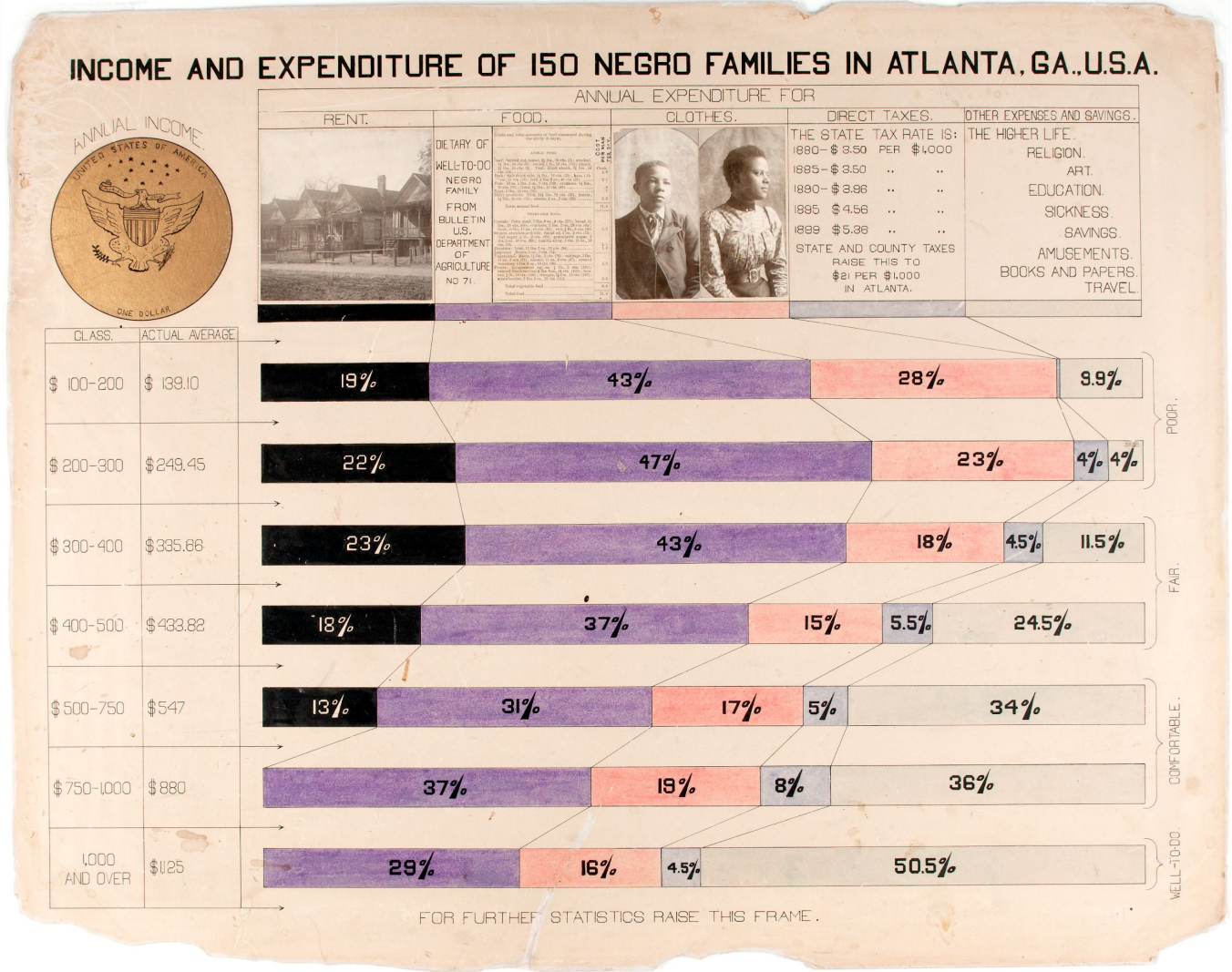
- `Table.read_table(filename)` - reads a table from a spreadsheet
- `Table()` - an empty table

- 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 – $3.98 .. ..
1895 $4.56 .. ..
1899 $5.38 .. ..
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
$1,125
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`

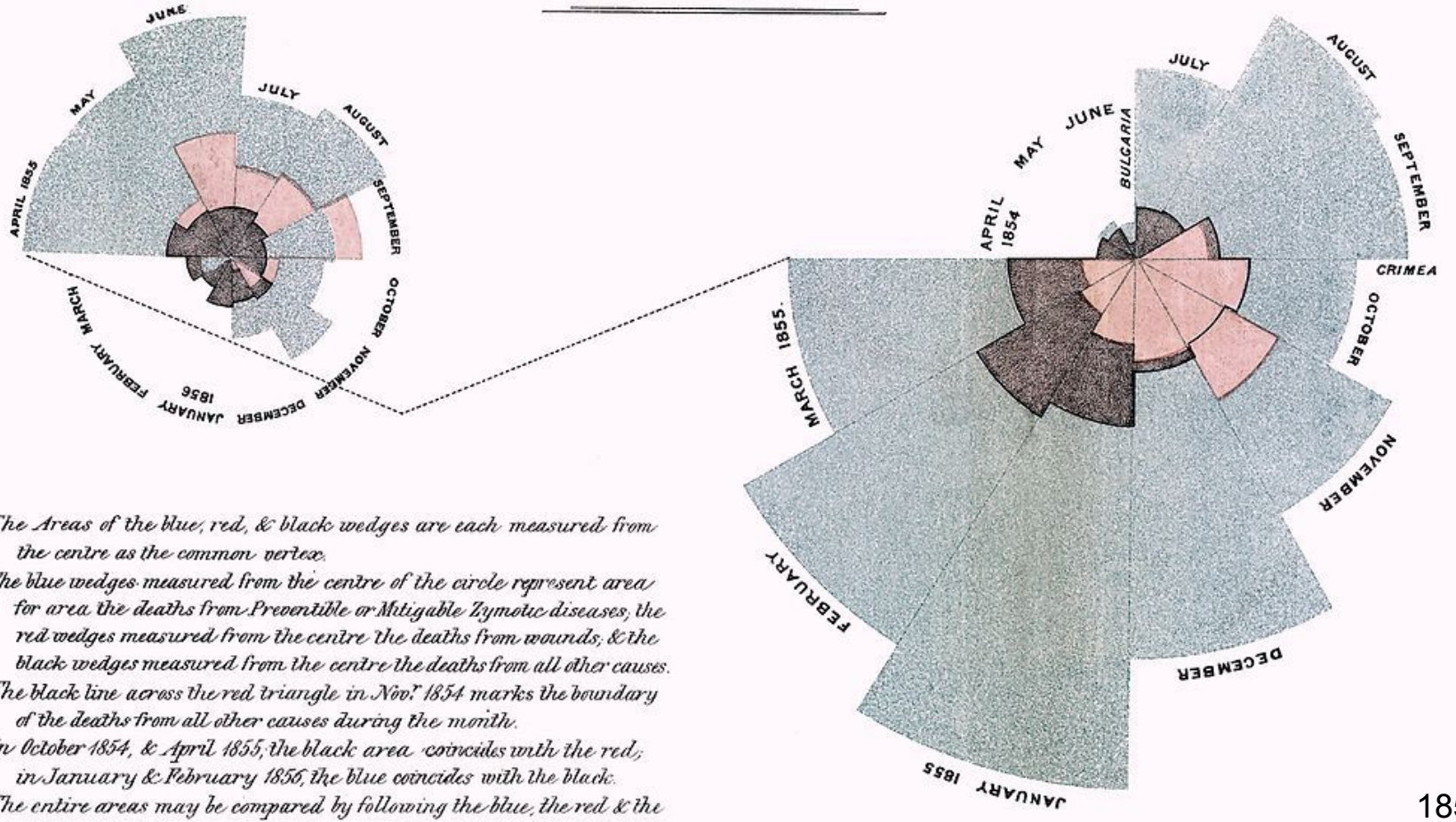
(Demo)

Another famous visualization

DIAGRAM of the CAUSES of MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 to MARCH 1856.

1.
APRIL 1854 to MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

1858

Bonus Example

Charles Joseph Minard, 1781-1870

- French civil engineer who created one of the greatest graphs of all time
- Visualized Napoleon's 1812 invasion of Russia, including
  - the number of soldiers
  - the direction of the march
  - latitude and longitude
  - temperature on the return journey
  - dates in November and December

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaient été détachés sur Minsk et Mobilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

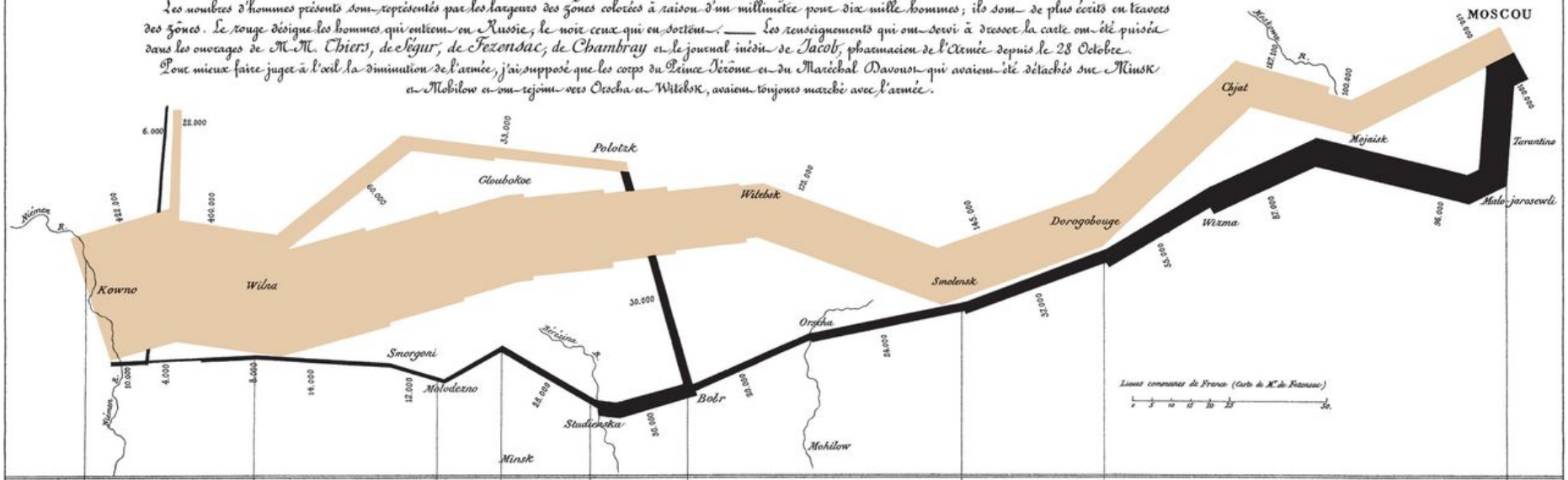
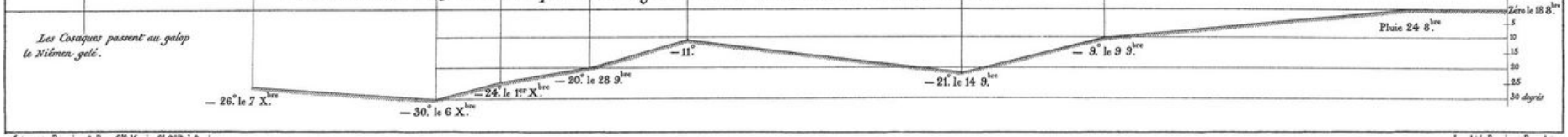


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



Autog. par Regnier, 8 Pas. St.te Marie St.t Gen.ve à Paris.

Imp. Lith. Regnier et Douradet.

Some of Minard's Data

Longitude	Latitude	City	Direction	Survivors
24	54.9	Kaunas	Advance	340000
30.2	55.2	Vitebsk	Advance	175000
32	54.8	Smolensk	Advance	145000
37.6	55.8	Moscow	Advance	100000
34.3	55.2	Vyazma	Retreat	55000
32	54.6	Smolensk	Retreat	24000
30.4	54.4	Orsha	Retreat	20000
26.8	54.3	Maladyechna	Retreat	12000
24.1	54.4	Kaunas	Retreat	4000

(Demo)

Discussion Question

Use the table functions we learned this week to find the southernmost city along the army's retreat.