

Lecture 7

Visualization, Randomness, and Computation

History of Data Science, Spring 2022 @ UC San Diego
Suraj Rampure

Announcements

- Look at others' Homework 6 websites!
 - Posted in the #homework6 channel on Slack.
 - If your images don't load, please fix them ASAP, and message if you need help.
- Homework 7 will be released by **noon tomorrow**.
- **Make sure to read Homework 5 solutions (posted on Campuswire)!**
 - Many misconceptions about the Law of Large Numbers.

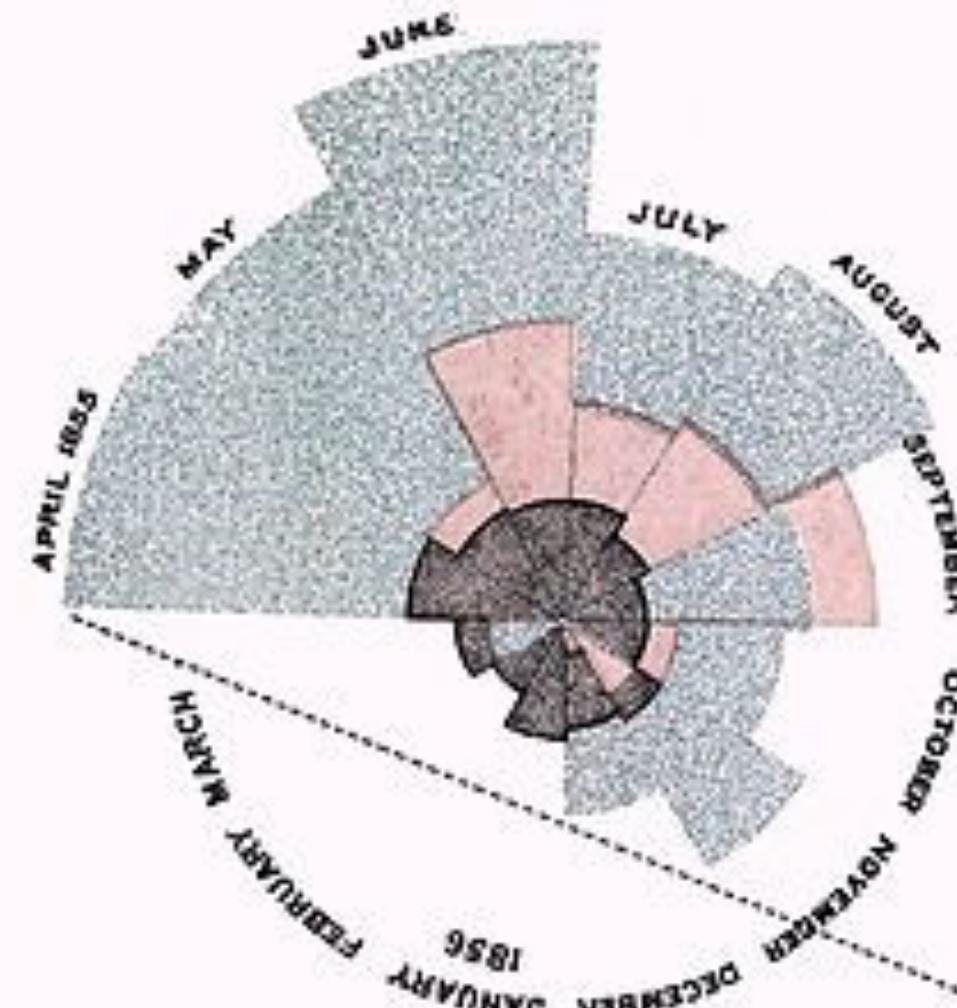
Agenda

- Finish visualization.
- Revisit the development of the Gaussian distribution.
- Abacus and the binary number system.

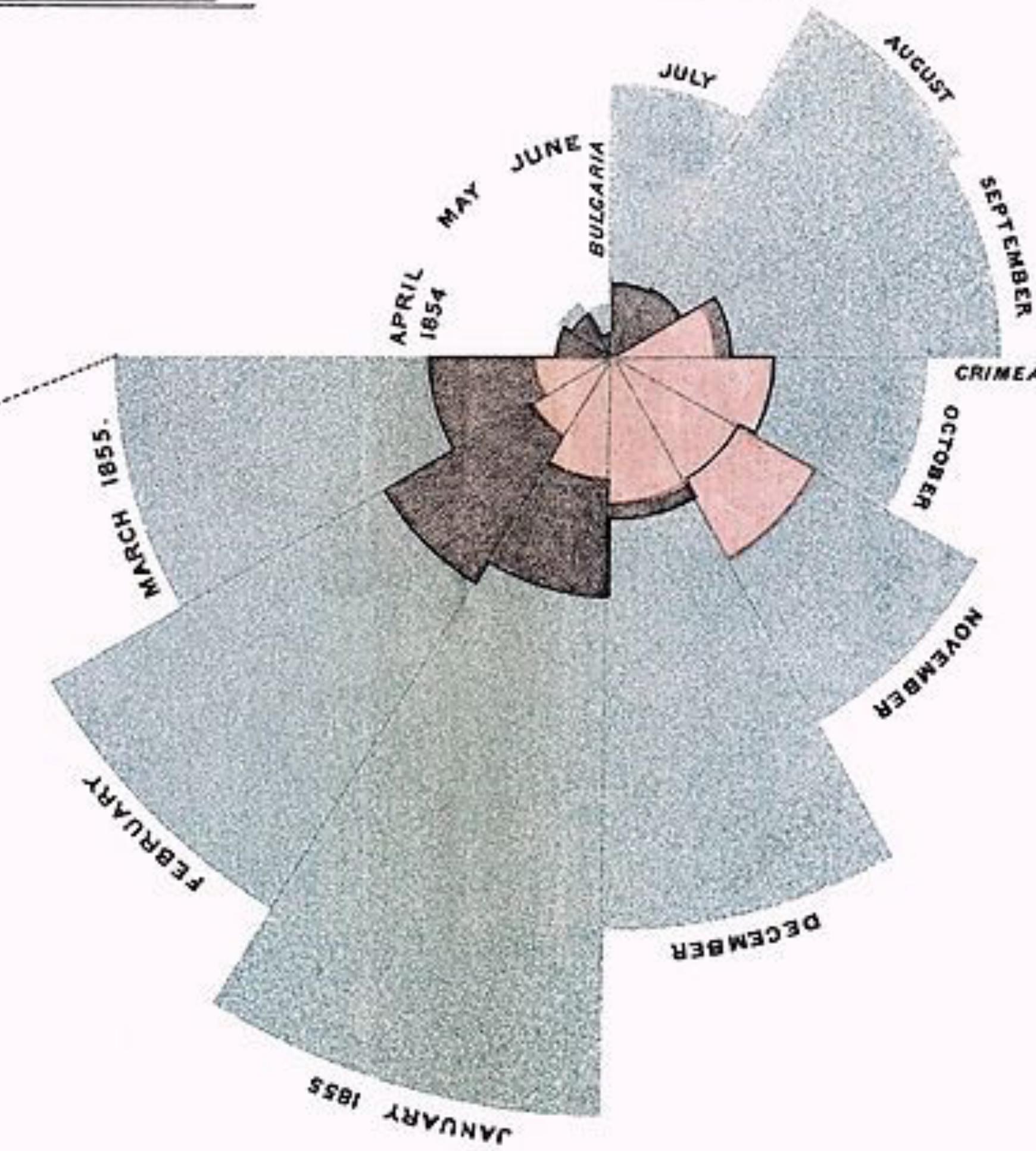
Visualization

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. 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.

1855: Florence Nightingale's depiction of the deaths of British soldiers in the Crimean war. Florence Nightingale is known as the founder of modern nursing.

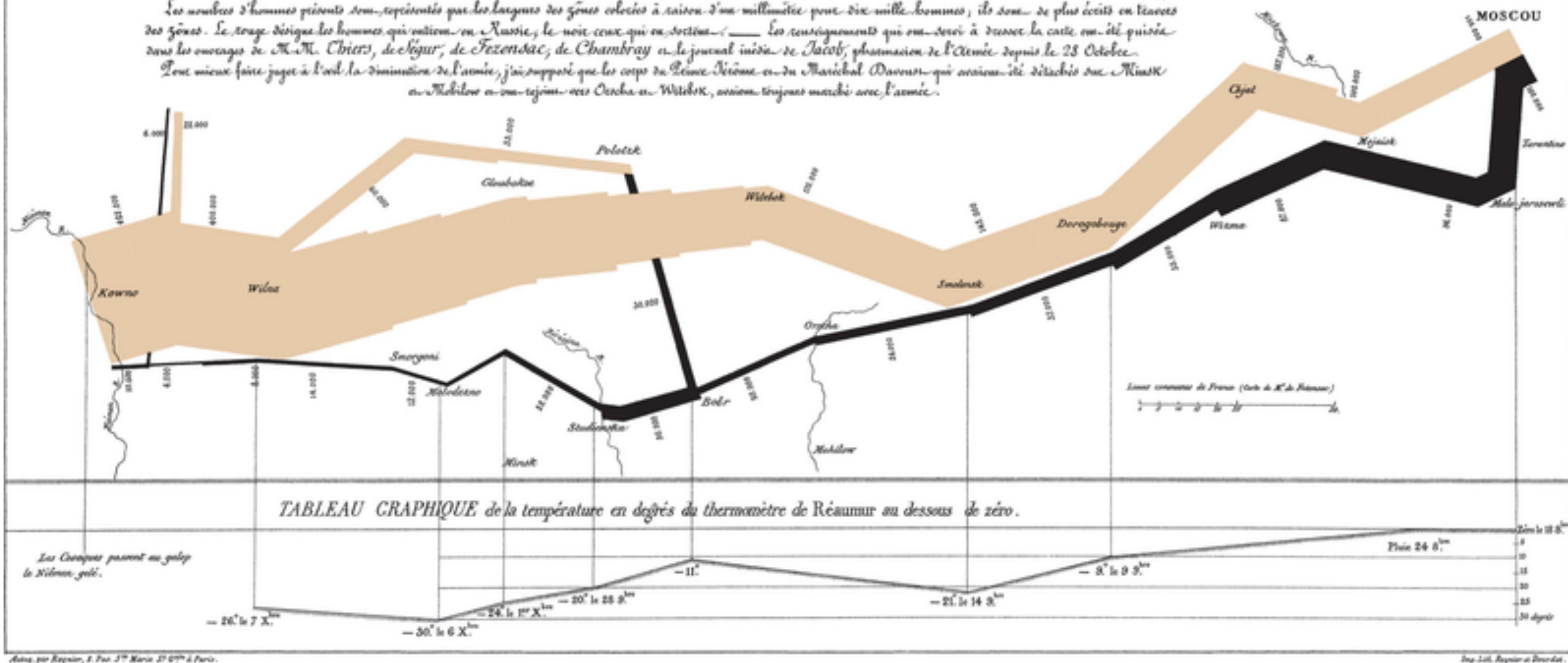
Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

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

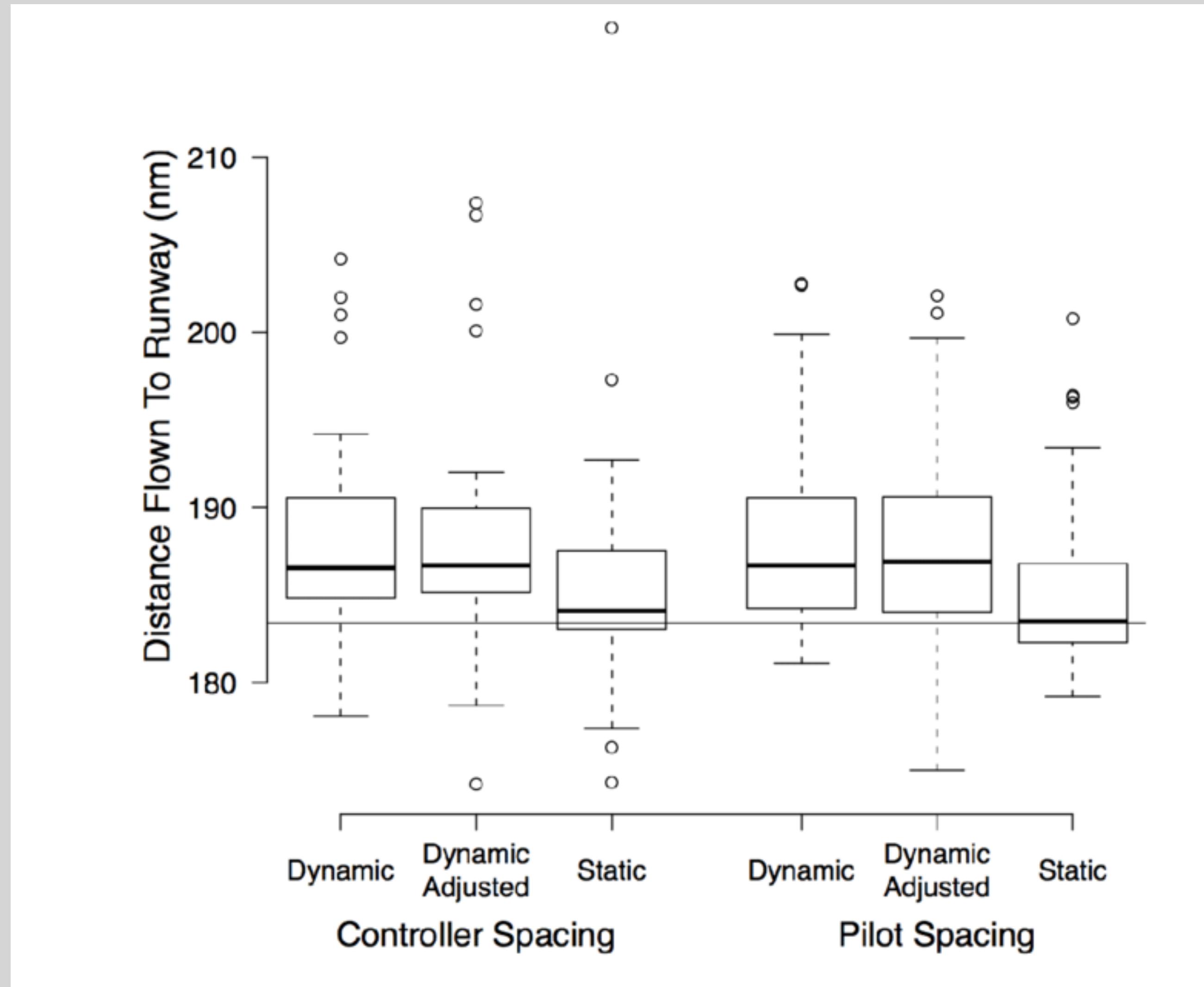
Paris, le 20 Novembre 1869.

Les nombres d'hommes présentés sont, représentés par les larges des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui restent en Russie, le noir ceux qui se retirent. — Les renseignements qui ont servi à dresser la carte ont été pris dans les ouvrages de M.-A. Chiers, de Clémier, de Fezensac, de Chambray et le journal intime de Jacob, publication de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai rapporté que les corps de l'Armée Néerlandaise du Maréchal Davout qui avaient été détruits sur Niensk et Malibolz au mois d'août, avec Orelia et Witsch, avaient toujours marché avec l'armée.



1869: Charles Joseph Minard's visualization of the French invasion of Russia (led by Napoleon).

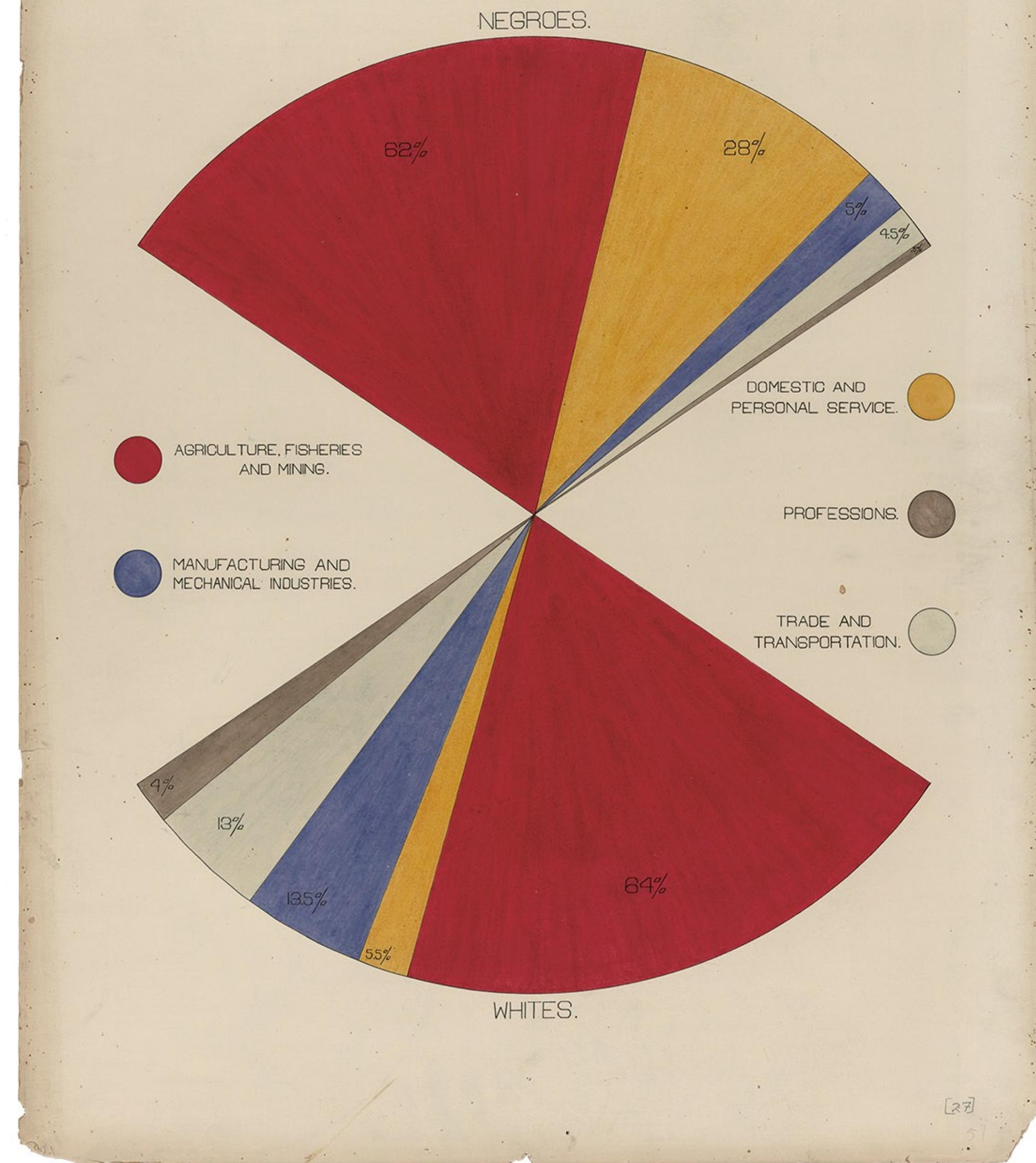


1973: John Tukey, who defined the term “Exploratory Data Analysis”, created the box plot, which describes a numerical distribution using a 5 number summary.

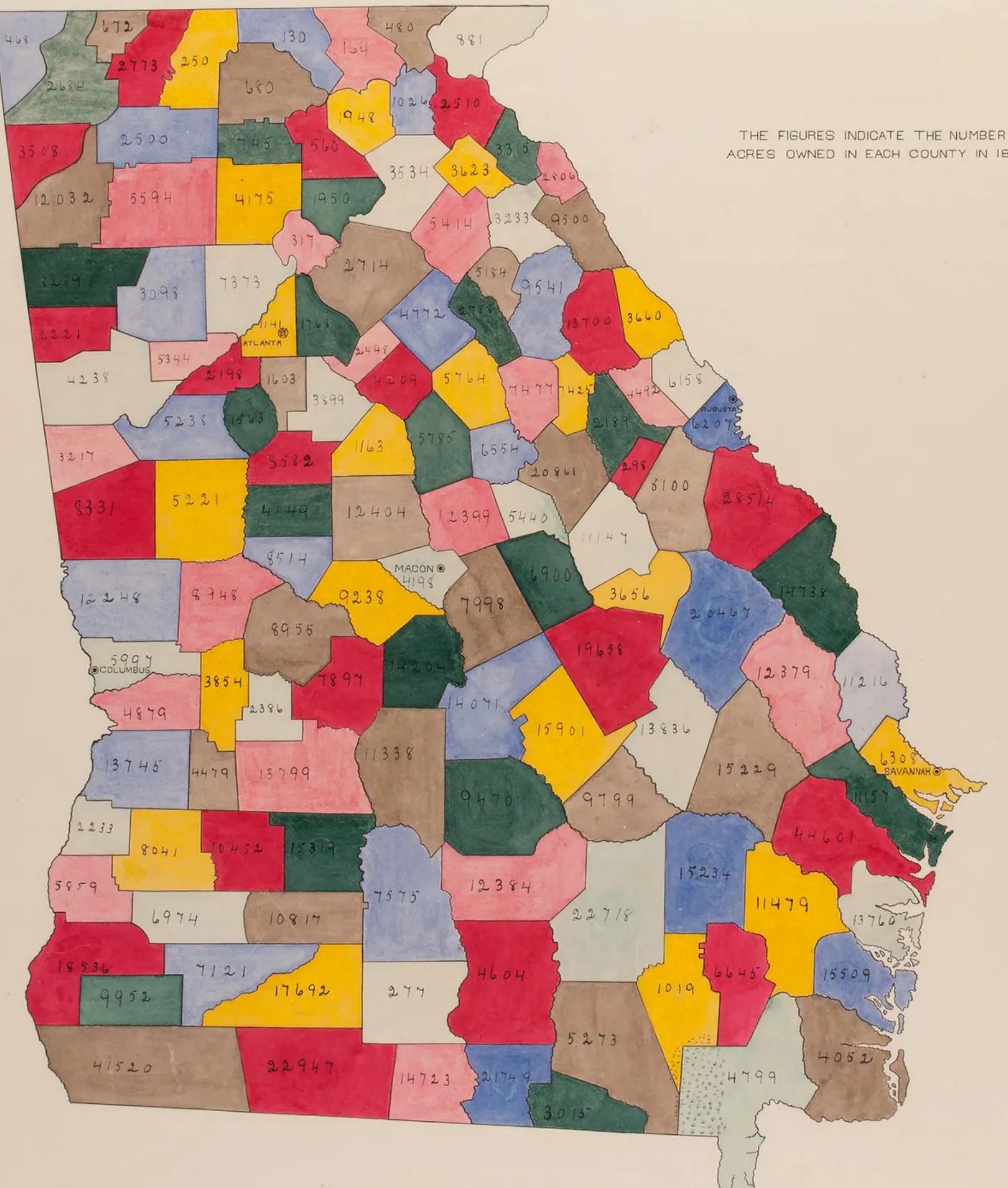
W. E. B. Dubois' visualizations of Black America

- W. E. B. Dubois (1868-1963) was an African-American historian, sociologist, and activist.
- He helped create the National Association for the Advancement of Colored People.
- He is also the first African-American to earn a PhD from Harvard.
- For the 1900 “Exposition Universelle” (world’s fair) in Paris, he created a display celebrating the advancements of African-Americans.
- Part of the display was a series of data visualizations (more in the reading).

OCCUPATIONS OF NEGROES AND WHITES IN GEORGIA.



LAND OWNED BY NEGROES IN GEORGIA, U.S.A. 1870-1900.



THE FIGURES INDICATE THE NUMBER OF ACRES OWNED IN EACH COUNTY IN 1899.

Derivation of the Gaussian distribution

Gauss and least squares

- **Recall from Lecture 4:** one of the key differences between the approaches to least squares by Gauss and Legendre was that Gauss linked the theory of least squares to probability theory.
- Specifically, he posed the least squares **model** where

$$y_i = a + bx_i + \epsilon_i$$

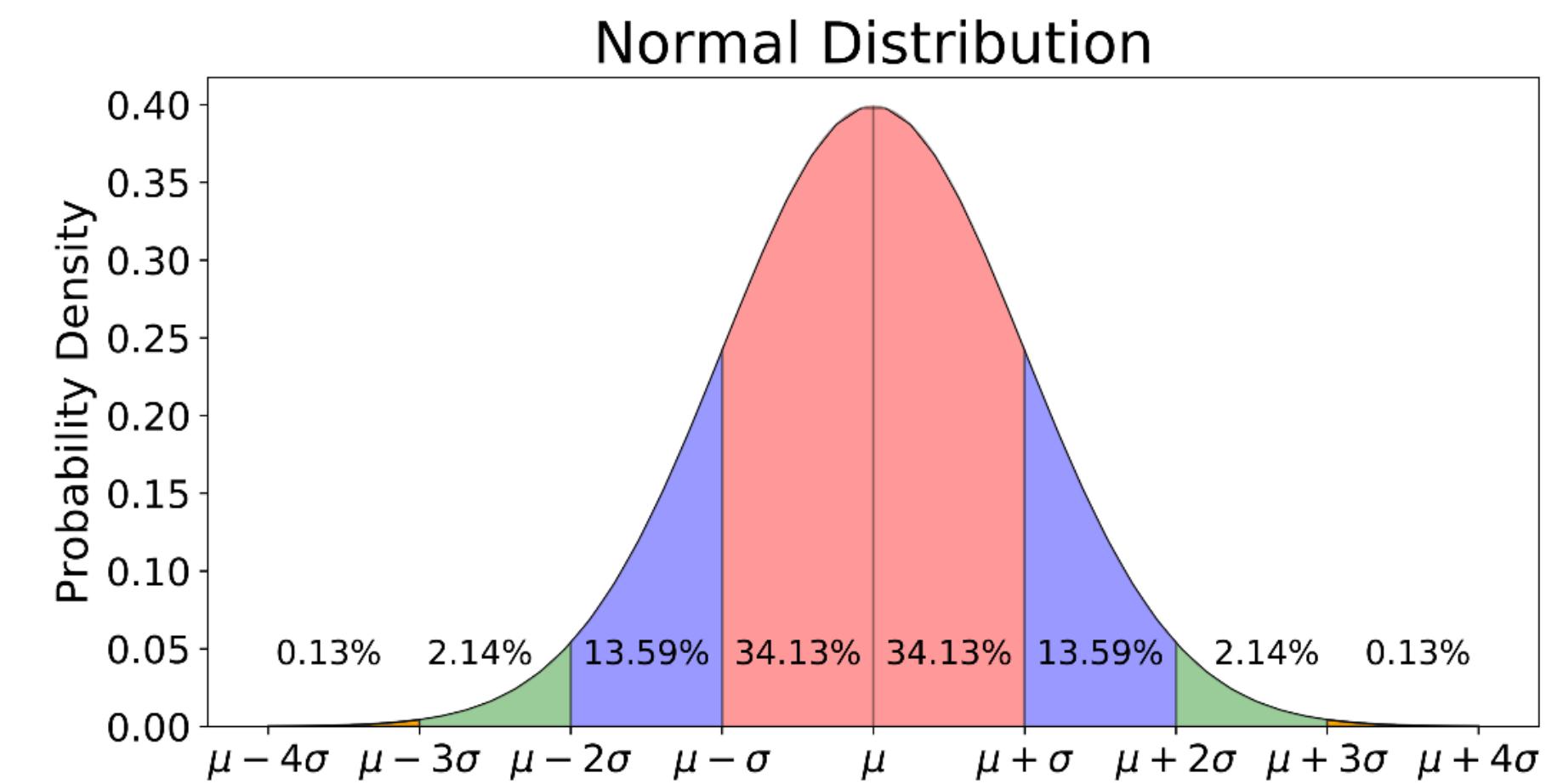
where ϵ_i is a **random variable** that follows the following **error distribution**:

$$\phi(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Where did this distribution come from? 🤔

Criteria for the Gaussian distribution

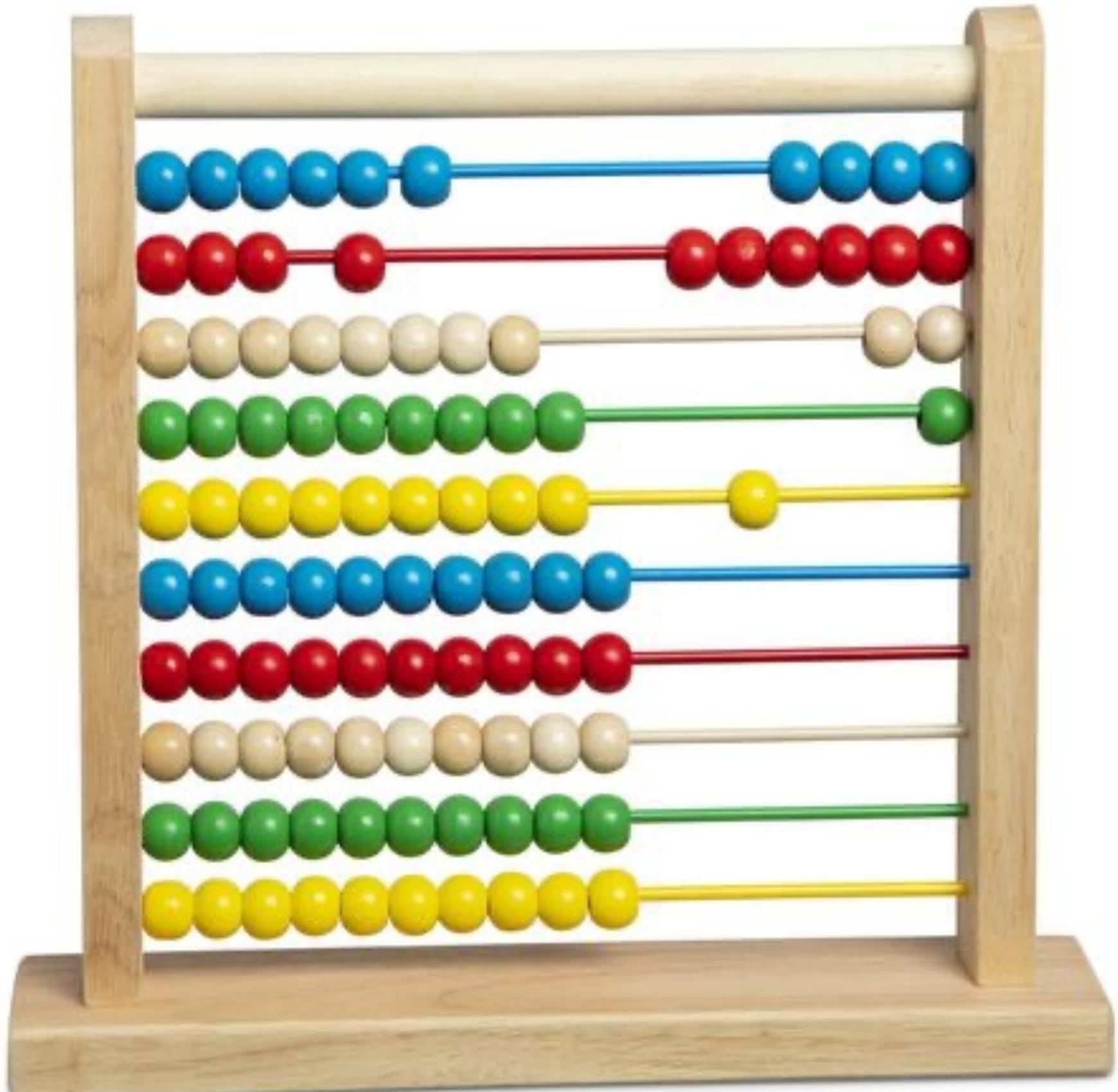
- Gauss described that the distribution of errors should satisfy three criteria:
 1. Small errors are more likely than large errors.
 2. For any real number ϵ the likelihood of errors of magnitude ϵ and $-\epsilon$ are equal.
 3. In the presence of several measurements of the same quantity, the most likely value of the quantity being measured is their average.
- We will now derive the Gaussian distribution using just these three criteria. (Buckle up!)



Origins of computation

Abacus

- The **abacus** is an early form of a calculator, used in Ancient Mesopotamia and China as early as 1900 BC¹.
 - Abacuses are still in use today!
 - It can be used to evaluate addition, subtraction, multiplication, and division without needing symbolic arithmetic.
 - The number system we use today is known as the Hindu-Arabic number system, and dates to ~700 AD.
 - See [this site](#) for a digital abacus.



1. <https://criticallyconsciouscomputing.org/history>

Computers

- In 1613¹, English poet Richard Braithwait published a book titled *The Yong Mans Gleanings*, which contains the earliest known reference to the word “computer”.
- In it, a “computer” was defined as a **person** who performed arithmetic and algebraic operations.
- “Computer”s referred to people as recent as the mid-1900s.
 - Then, most computers were women, because they could be paid less than men.

1. <https://criticallyconsciouscomputing.org/history>

Leibniz and binary

- Gottfried Wilhelm Leibniz (1646-1716) – one of the “creators of calculus” that we studied earlier in the course – is also credited for developing the **binary number system**.
 - There are two digits in binary – 1 and 0. “Bit” stands for “binary digit.”
- At the time, he had no practical use for it, but in modern computing all code is represented at a low-level in binary.
 - Transistors – the building block of computer processors – are like switches that can either be “on” (1) or “off” (0).
 - One of this week’s readings contains a translation of his original work that discussed binary numbers, titled *“EXPLANATION OF BINARY ARITHMETIC, WHICH USES ONLY THE CHARACTERS 0 AND 1, WITH SOME REMARKS ON ITS USEFULNESS, AND ON THE LIGHT IT THROWS ON THE ANCIENT CHINESE FIGURES OF FU XI”*.

Number systems

- Our standard number system, base 10, has 10 digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9.

- We can decompose the number 3249_{10} as follows:

$$3249_{10} = 3 \cdot 10^3 + 4 \cdot 10^2 + 2 \cdot 10^1 + 9 \cdot 10^0$$

- Similarly, base 2 (i.e. binary) only has 2 digits: 0 and 1.
 - To convert from binary to base 10, we can follow a similar procedure.

$$1001_{10} = 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 8 + 1 = 9$$

Converting between binary and base 10

Example: Convert 324_{10} to binary.

Example: Convert 110001011_2 to base 10.

Binary arithmetic

- Arithmetic in binary largely works the same way that arithmetic in base 10 works.
- **Example:** multiply 10110_2 and 11_2 .

Symbols

- We've looked at how base 10 numbers can be stored in binary.
- But base 10 numbers are not the only thing our computers need to store and work with.
 - What about negative numbers? Decimals?
 - Strings?
 - Colors?
 - All of these can be stored in binary as well.



Boolean algebra

- George Boole (1815-1864) was an English mathematician.
- In 1854, he published *An investigation into the Laws of Thought, on Which are founded the Mathematical Theories of Logic and Probabilities*, in which he laid the foundations of **Boolean algebra**.
- In Boolean algebra, there are two values – 1 (True) and 0 (False), and three operators – AND, OR, and NOT.
- Fun fact: at the age of 19, Boole created his own elementary school!

1st. Disjunctive Syllogism.

$$\begin{array}{l} \text{Either } X \text{ is true, or } Y \text{ is true (exclusive),} \\ \text{But } X \text{ is true,} \\ \text{Therefore } Y \text{ is not true,} \end{array} \quad \frac{x + y = 2xy = 1}{\begin{array}{l} x = 1 \\ \therefore y = 0 \end{array}}$$

$$\begin{array}{l} \text{Either } X \text{ is true, or } Y \text{ is true (not exclusive),} \\ \text{But } X \text{ is not true,} \\ \text{Therefore } Y \text{ is true,} \end{array} \quad \frac{\begin{array}{l} x + y = xy = 1 \\ x = 0 \\ \therefore y = 1 \end{array}}{}$$

2nd. Constructive Conditional Syllogism.

$$\begin{array}{l} \text{If } X \text{ is true, } Y \text{ is true,} \\ \text{But } X \text{ is true,} \\ \text{Therefore } Y \text{ is true,} \end{array} \quad \frac{\begin{array}{l} x(1-y) = 0 \\ x = 1 \\ \therefore 1-y = 0 \text{ or } y = 1. \end{array}}{}$$

3rd. Destructive Conditional Syllogism.

$$\begin{array}{l} \text{If } X \text{ is true, } Y \text{ is true,} \\ \text{But } Y \text{ is not true,} \\ \text{Therefore } X \text{ is not true,} \end{array} \quad \frac{\begin{array}{l} x(1-y) = 0 \\ y = 0 \\ \therefore x = 0 \end{array}}{}$$

4th. Simple Constructive Dilemma, the minor premiss exclusive.

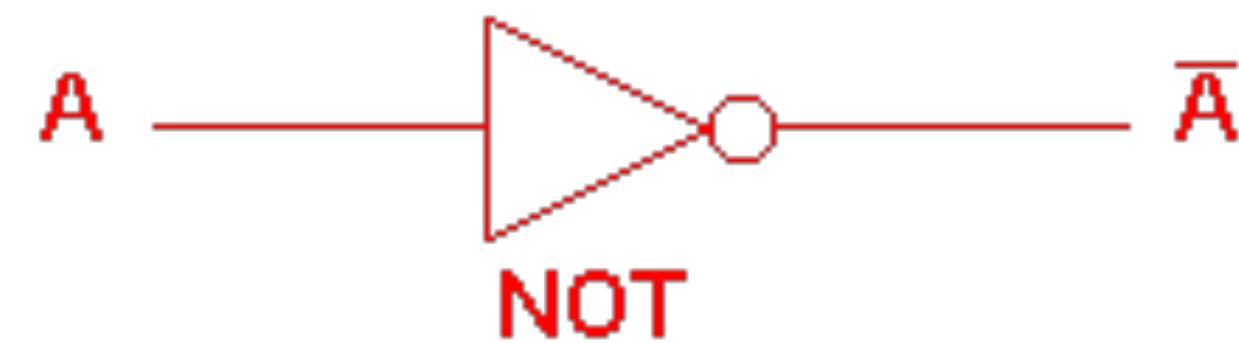
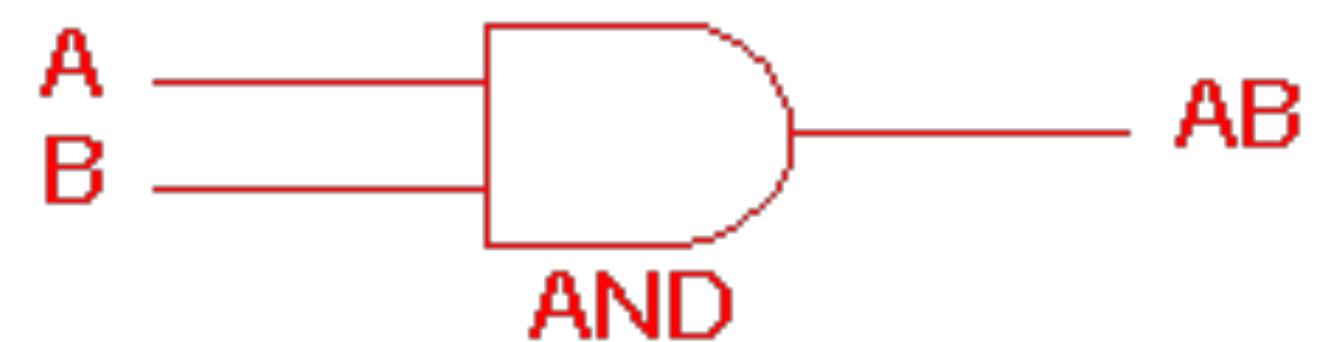
$$\begin{array}{l} \text{If } X \text{ is true, } Y \text{ is true,} \\ \text{If } Z \text{ is true, } Y \text{ is true,} \\ \text{But Either } X \text{ is true, or } Z \text{ is true,} \end{array} \quad \frac{\begin{array}{l} x(1-y) = 0, \quad (41), \\ z(1-y) = 0, \quad (42), \\ x + z - 2xz = 1, \quad (43). \end{array}}{}$$

From the equations (41), (42), (43), we have to eliminate x and z . In whatever way we effect this, the result is

$$\begin{array}{l} y = 1; \\ \text{whence it appears that the Proposition } Y \text{ is true.} \end{array}$$

Boolean algebra

- In Boolean algebra, there are two values – 1 (True) and 0 (False), and three operators – AND, OR, and NOT.
- All other operations can be constructed using a combination of these three operators.
- Circuits use Boolean algebra to control the flow of current.



That's all!