STAT 184: Introduction to R

Padma Tanikella
Dept. of Statistics, Penn State

Getting to know each other

- Your name, the year and department you belong to.
- Any prior data analysis experience using calculators, Excel sheets etc.?
- Any prior programming experience?

Textbooks and syllabus

- We have four recommended textbooks, all of which are freely available. (Check Canvas course homepage.)

- Daily agenda Excel sheet and syllabus page are available on the course homepage.

- The Daily agenda sheet has recommended readings associated for each day of class.

Regular R Activities

- We will have regular R activities associated with lectures.
- All activities will be posted and graded on Canvas.
- Each R activity is due on the third day of classes following the time it is first released.

Midterm and Final Project

- Midterm will be held during class timings on 19th July.
- Midterm will be a closed-book pencil-code exam.

- Final project will have three different parts:
  - Submit a topic idea (due 25 July, 11:59pm)
  - Submit initial exploratory data analysis (EDA) (due 31 July. 11:59pm)
  - Submit a final report (due 7 August, 11:59pm)
- Further details about the Midterm and Final project will eventually be posted on Canvas.

Office hours and TA

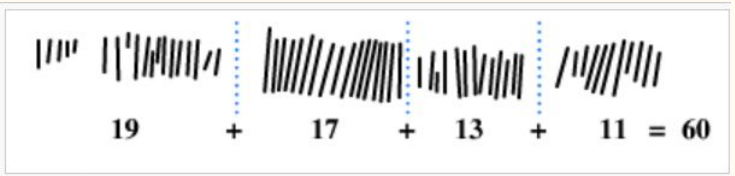
- Haochen Zhang is the course TA.
- My PSU email ID is pbt5137@psu.edu and Haochen's PSU email ID is hqz5340@psu.edu
- Do not use the Canvas discussion board to communicate personal information.
- Information about office hours will soon be posted on the Canvas homepage.

Should I learn to program?

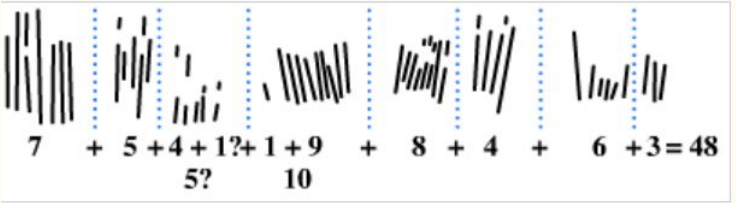
Let us look at how the nature of data has changed over time.

Ishango Bone (~20,000 BCE)

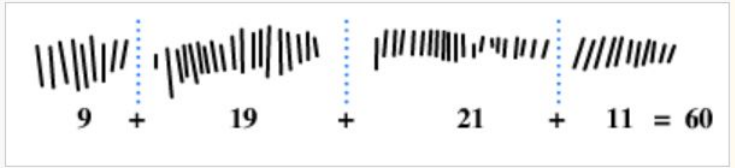
This image shows both the front and back of the Ishango bone.



Left column



Centre column



Right column

Reference: Ishango Bone, 2007 Schools Wikipedia Selection

Counting tables and Abacus

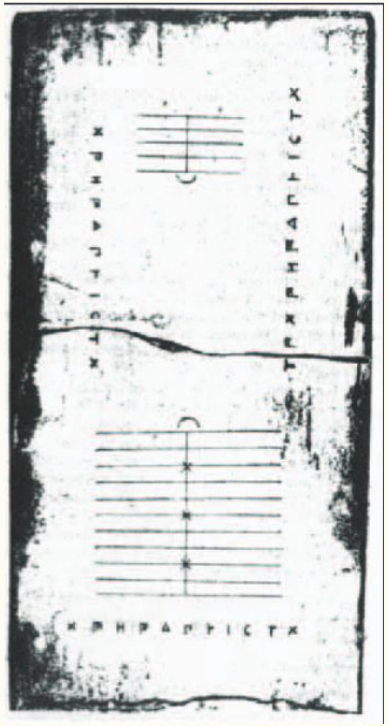


Fig 1: Salamis Counting board

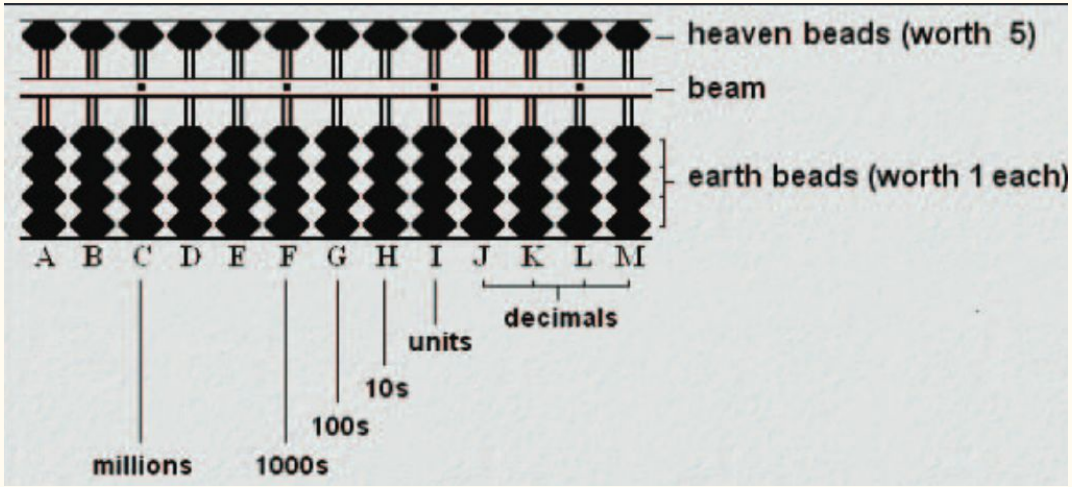
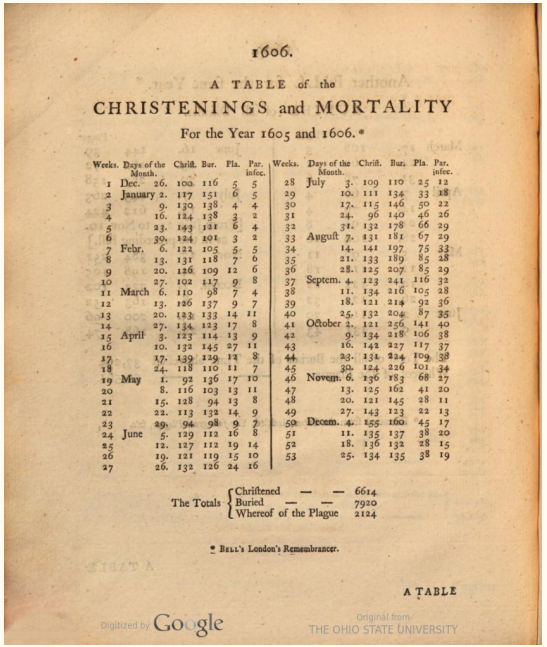


Fig. 2: Soroban Abacus

Reference: *The History of the Abacus*, Kevin Samoly, Ohio Journal of School Mathematics, Number 65 (Spring 2012)

John Graunt's work in demography (1662)



1606.

A TABLE of the

CHRISTENINGS and MORTALITY

For the Year 1605 and 1606. *

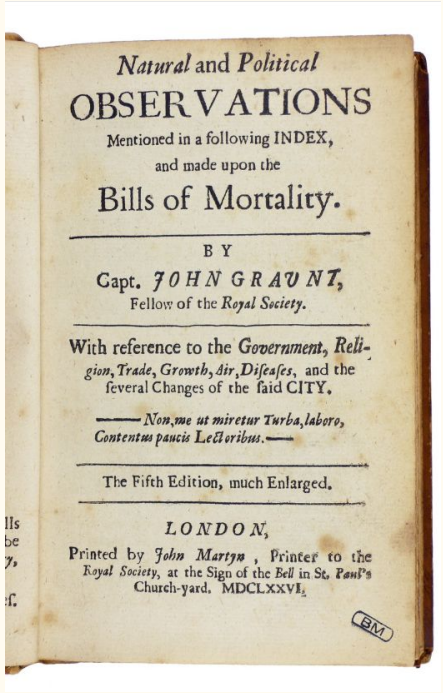
Weeks.	Days of the Month.	Christ.	Bur.	Fla.	Par. infec.	Weeks.	Days of the Month.	Christ.	Bur.	Fla.	Par. infec.
1	Dec. 26.	100	116	5	5	28	July 3.	109	110	25	12
2	January 2.	117	151	6	5	29	10.	111	134	33	18
3	9.	130	138	4	4	30	17.	115	146	50	22
4	16.	124	138	3	2	31	24.	96	140	46	26
5	23.	143	121	6	4	32	31.	132	178	66	29
6	30.	124	101	3	2	33	August 7.	131	181	67	29
7	Febr. 6.	122	105	5	5	34	14.	141	197	75	33
8	13.	131	118	7	6	35	21.	133	189	85	28
9	20.	126	109	12	6	36	28.	125	207	85	29
10	27.	102	117	9	8	37	Septem. 4.	123	241	116	32
11	March 6.	110	98	7	4	38	11.	134	216	105	28
12	13.	126	137	9	7	39	18.	121	214	92	36
13	20.	133	133	14	11	40	25.	132	204	87	35
14	27.	134	123	17	8	41	October 2.	121	256	141	40
15	April 3.	123	114	13	9	42	9.	134	218	106	38
16	10.	132	145	27	11	43	16.	142	227	117	37
17	17.	139	129	12	8	44	23.	131	224	109	38
18	24.	118	110	11	7	45	30.	124	226	101	34
19	May 1.	92	136	17	10	46	Novem. 6.	136	183	68	27
20	8.	116	103	13	11	47	13.	125	162	41	20
21	15.	128	94	13	8	48	20.	121	145	28	11
22	22.	113	132	14	9	49	27.	143	123	22	13
23	29.	94	98	9	7	50	Decem. 4.	155	160	45	17
24	June 5.	129	112	16	8	51	11.	135	137	38	20
25	12.	127	112	19	14	52	18.	136	132	28	15
26	19.	121	119	15	10	53	25.	134	135	38	19
27	26.	132	126	24	16						

The Totals { Christened — — 6614. Buried — — 7920. Whereof of the Plague 2124.

* Bell's London's Remembrancer.

A TABLE

Fig 1: Bill of Mortality for the year 1605-1606



Natural and *Political*

OBSERVATIONS

Mentioned in a following INDEX,

and made upon the

Bills of Mortality.

BY

Capt. JOHN GRAUNT,

Fellow of the *Royal Society*.

With reference to the *Government, Religion, Trade, Growth, Air, Diseases*, and the several Changes of the said CITY.

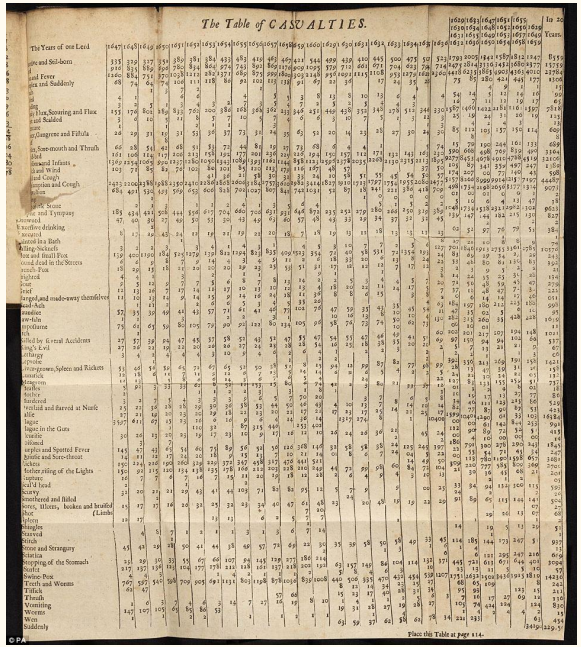
Non, me ut miretur Turba, laboro,
Contentus paucis Lectoribus.

The Fifth Edition, much Enlarged.

LONDON,

Printed by *John Martyn*, Printer to the *Royal Society*, at the Sign of the *Bell* in St. *Paul's* Church-yard. MDCLXXVI.

Fig 2: Title page of the 5th edition of Graunt's book.



The Table of CASUALTIES.

Fig 3: "Table of Casualties" from Graunt's book.

Reference: Wikipedia page on John Graunt

Herman Hollerith invents an electronic tabulating machine (1880)

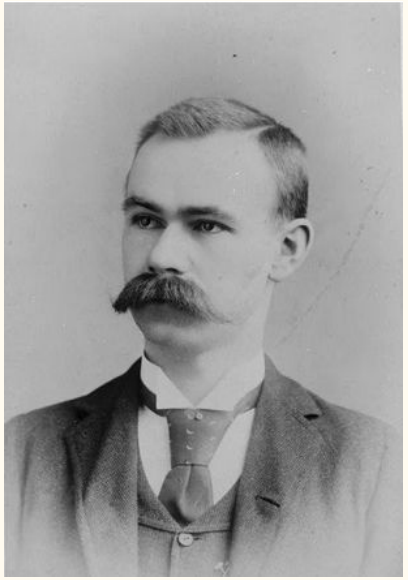


Fig 1. Herman Hollerith

(Image taken from US Census Bureau article on Herman Hollerith)

*"Three contestants accepted the Census Bureau's challenge. The first two contestants captured the data in **144.5 hours** and **100.5 hours**. The third contestant, Herman Hollerith, completed the data capture process in **72.5 hours**."*

Reference: The Hollerith Machine, US Census Bureau article.

Fig 2: Hollerith's electronic tabulator

During the 1900s...

- 1965 - US Government plans the world's first data center to store 742 million tax returns and 175 million sets of fingerprints on magnetic tape.
- 1989 - Possibly the first use of the term Big Data (without capitalization) in the way it is used today, by international best-selling author Erik Larson.
- 1991 - Computer scientist Tim Berners-Lee announced the birth of what would become the Internet as we know it today.
- 1997 - Google search engine debuts this year!
- 1999 - Possibly first use of the term "Internet of Things", to describe the growing number of devices online and the potential for them to communicate with each other, often without a human "middle man".

Reference: World Economic Form article "A brief history of big data everyone should read (2015) "

What are the different kinds of data collected these days?