DATA SCIENCE IN CONTEXT
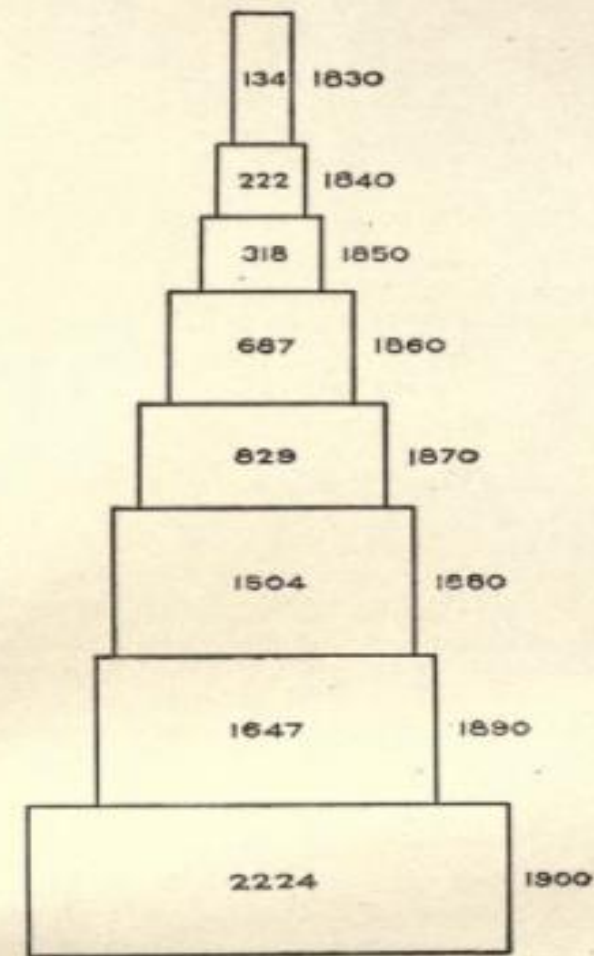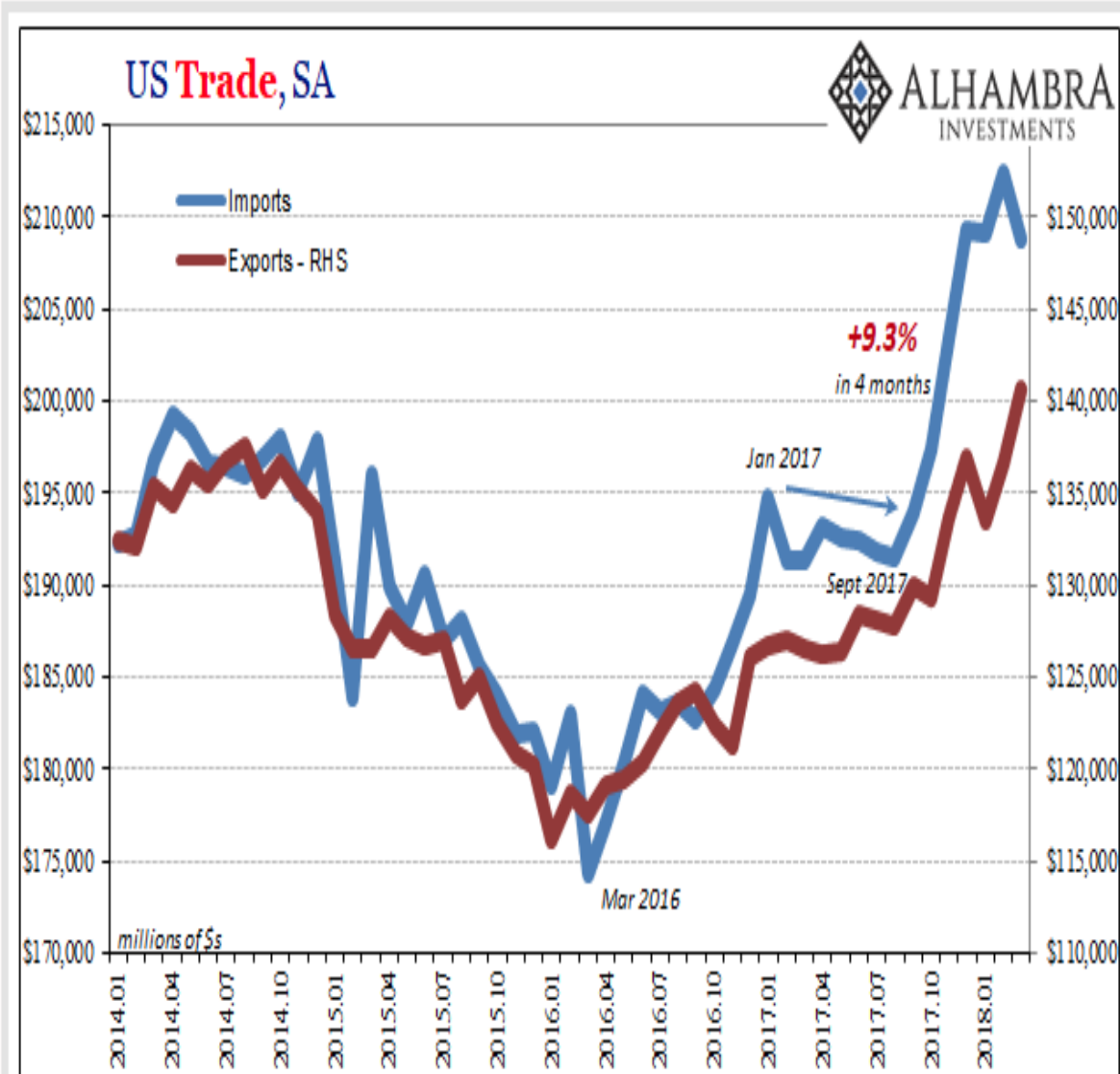
A brief history of data storage-analysis

Presented by **Alexis Mekueko**

Github link: https://github.com/asmozo24/Data-Science-in-Context

Contrast of Data Visualization 1830-2018

US Trade Balance, Jan 2014 – May 2018

US **Trade**, SA

ALHAMBRA INVESTMENTS

— Imports
— Exports - RHS

+9.3%
in 4 months

Jan 2017

Sept 2017

Mar 2016

millions of $s

134   1830
222   1840
318   1850
687   1860
829   1870
1504 1880
1647 1890
2224 1900

The Philadelphia Commercial Museum

Fig. 66.   Commerce of the United States since 1830.   The Sum of Annual Exports and Imports.   Values are given in Millions of Dollars

Such a chart as this is worse than none.   There is no scale in either direction.   The block for 1830 is drawn with a larger area than the block for 1840 which represents a larger quantity.   Compare with Fig. 67 and Fig. 68

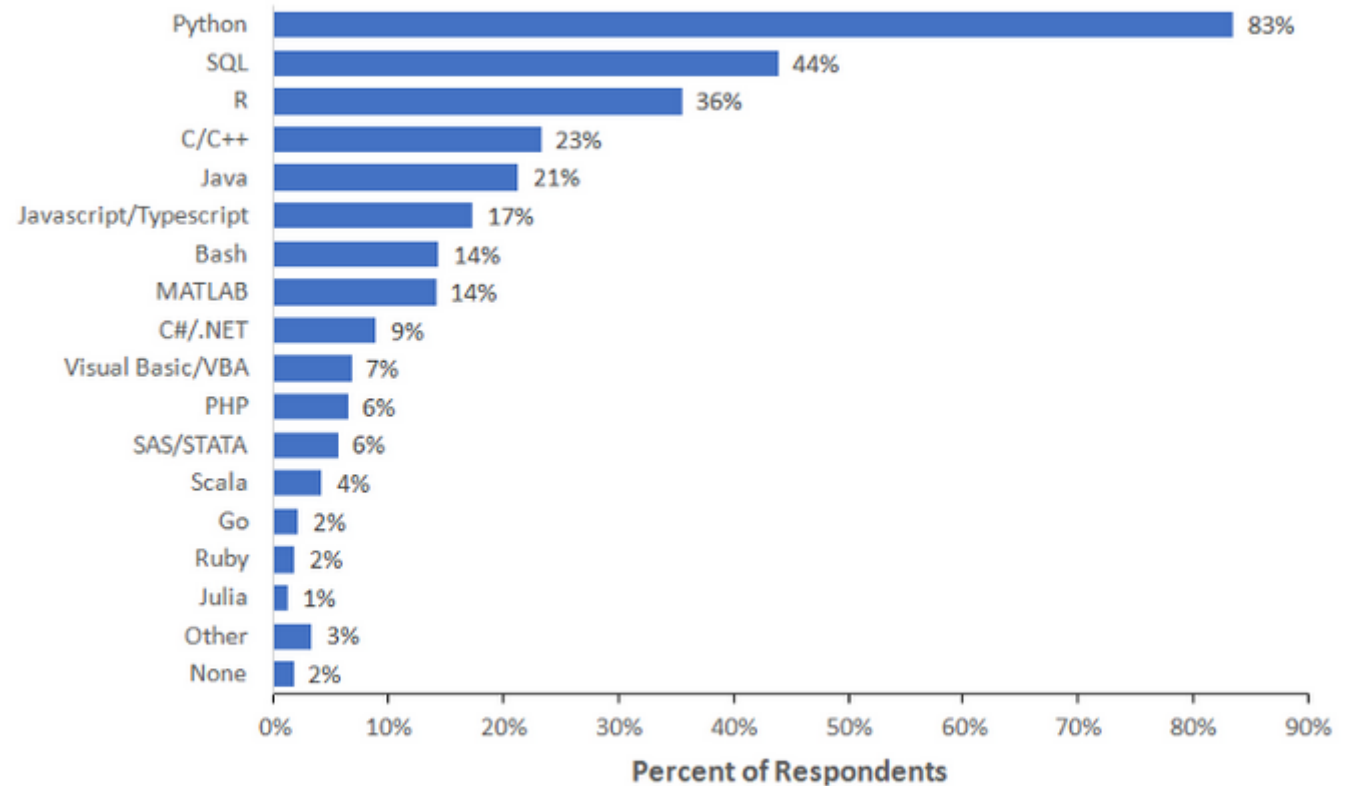# Data as Qualitative and Quantitative information

- 1830-Willard Cope Brinton, a consulting engineer and a pioneer in data visualization. In 1914 he published "Graphic methods for presenting facts"

- 1943- the ENIAC computing system was built by John Mauchly and J. Presper Eckert at the Moore School of Electrical Engineering of the University of Pennsylvania.

- Random Access Memory (RAM) In 1948 by Professor Fredrick Williams, and colleagues.

- 1958- Jack S. Kilby and later Robert Noyce were pioneers for the invention of the Integrated Circuit.

- In 1960, Charles W. Bachman designed the Integrated Database System, the "first" DBMS. IBM,

- 1962- John Tukey published "The Future of Data Analysis." in the The Annals of Mathematical Statistics.

- 1974 - Peter Naur published the "Concise Survey of Computer Methods", 1st time the term "Data Science" is defined: "The science of dealing with data, once they have been established, while the relation of the data to what they represent is delegated to other fields and sciences."

- 1989, the first Knowledge Discovery in Databases (KDD) workshop was organized (core of the process is -finding knowledge in data).

- 1997 - 2005- from Michael Cox and David Ellsworth published "Application-controlled demand paging for out-of-core visualization" for the IEEE 8th conference on visualization then, Roger Mougalas from O'Reilly Media extended big data to Web 2.0 ...later Hadoop was created by Yahoo!

- From 2010 to 2020 data science has been growing at fast rate….

How is data used today?

- Companies use data to: make projection, help in decision making, get a roadmap, analyze past issues, research for innovation in nearly all domain in the economy.

- Programming languages:

- [Mackenzie Mitchell](#)

- Apps for data scientists:

- **Power BI- Tableau**

- IBM Cognos- Qlik- DataEye

- Data Counter Widget- Glasswire

- Data Monitor – My Data Manger

- Airdroid- MiXplorer Silver

- Resilio Sync-Solid Explorer- SAP BusinessObjects

- Roambi - Salesforce Wave - SAS Mobile BI

- Oracle Business Intelligence – Splunk

- Google Analytics

What programming language do you use on a regular basis?

Language	Percent of Respondents
Python	83%
SQL	44%
R	36%
C/C++	23%
Java	21%
Javascript/Typescript	17%
Bash	14%
MATLAB	14%
C#/.NET	9%
Visual Basic/VBA	7%
PHP	6%
SAS/STATA	6%
Scala	4%
Go	2%
Ruby	2%
Julia	1%
Other	3%
None	2%

Note: Data are from the 2018 Kaggle Machine Learning and Data Science Survey. You can learn more about the study here: http://www.kaggle.com/kaggle/kaggle-survey-2018. A total of 18827 respondents answered the question.

The Future is Zettabyte

- As AI and Machine Learning momentum is rapidly spreading in technology, much more resources would be needed.

- Cloud

- Big data center

- Powerful applications

- My recommendation: get certified in one application of your choice

- [Exam 70-778: Analyzing and Visualizing Data with Microsoft Power BI](#)

- [Exam DA-100: Analyzing Data with Microsoft Power BI](#)

- IDC predicts that the Global Datasphere will grow from 175 Zettabytes by 2025

References

- Graphic methods for presenting facts by Brinton, Willard Cope, 1914

- https://archive.org/details/graphicmethodsfo00brinrich/page/70/mode/2up

- US Trade Balance, Jan 2014 – May 2018

- https://snbchf.com/2018/05/snider-watching-imports/abook-may-2018-usexim-sa-png/

- **Beginner's Guide to the History of Data Science**

- https://dataconomy.com/2016/03/beginners-guide-history-data-science/

- https://www.kdd.org/kdd2015/

- **A Brief History of Database Management**

- https://www.dataversity.net/brief-history-database-management/

- **History of Computers: A Brief Timeline**

- https://www.livescience.com/20718-computer-history.html

- **A Very Short History Of Big Data**

- https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/?sh=6594d8965a18

- The Digitization of the World From Edge to Core

- https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

- **How Big is Big Data?**

- **https://towardsdatascience.com/how-big-is-big-data-3fb14d5351ba**

- **Programming Languages For Data Scientists**

- https://towardsdatascience.com/programming-languages-for-data-scientists-afde2eaf5cc5

- **Power BI certification:**

- **https://docs.microsoft.com/en-us/learn/certifications/exams/70-778**

- **https://docs.microsoft.com/en-us/learn/certifications/exams/da-100**