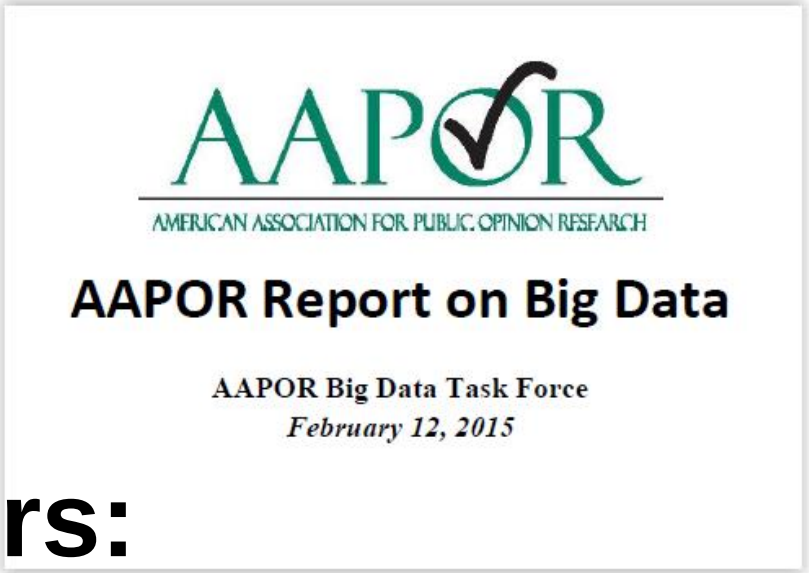
The Big Data team

Marcus Berg



Task Force Members:

***Lilli Japec**, Co-Chair, Statistics Sweden*

***Frauke Kreuter**, Co-Chair, JPSM at the U. of Maryland, U. of Mannheim & IAB*

***Marcus Berg**, Stockholm University*

***Paul Biemer**, RTI International*

***Paul Decker**, Mathematica Policy Research*

***Cliff Lampe**, School of Information at the University of Michigan*

***Julia Lane**, American Institutes for Research*

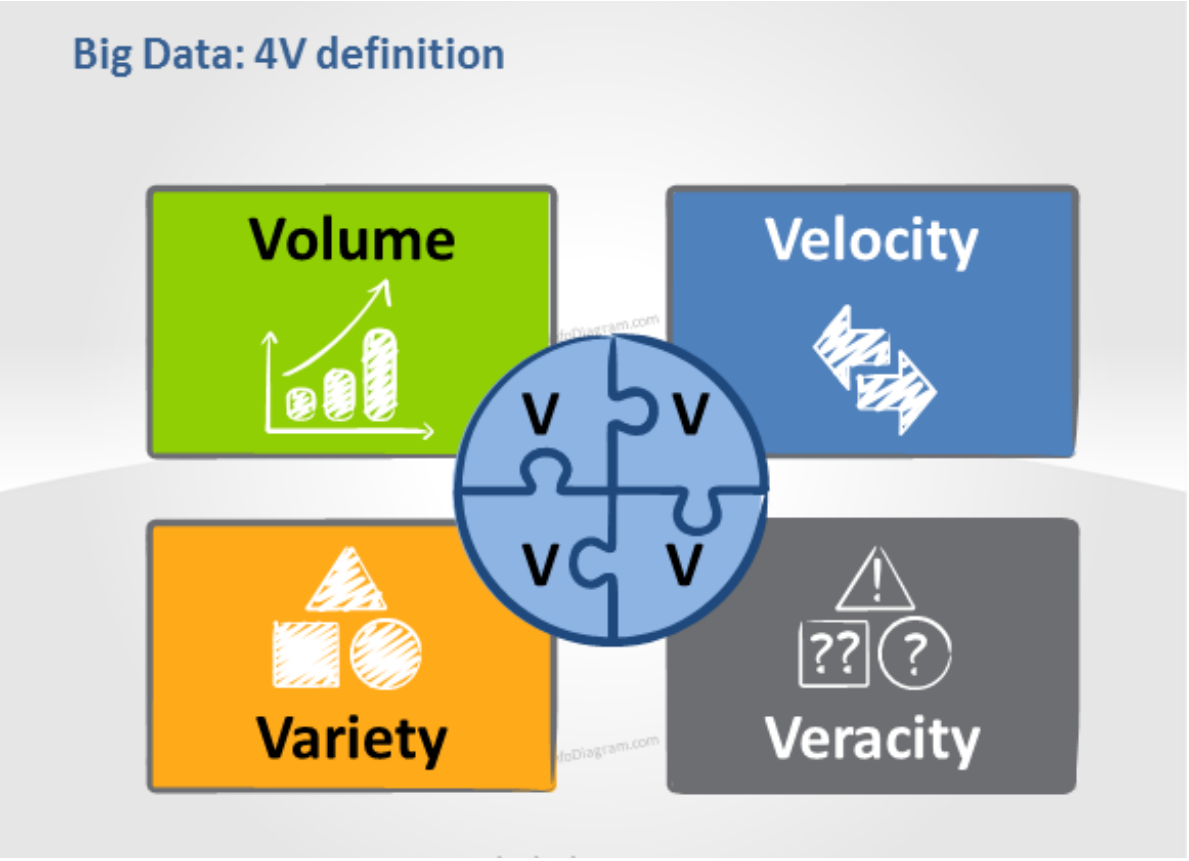
***Cathy O'Neil**, Johnson Research Labs*

***Abe Usher**, HumanGeo Group*

Acknowledgement: We are grateful for comments, feedback and editorial help from Eran Ben-Porath, Jason McMillan, and the AAPOR council members.

Another way to visualize it

because visualization is important

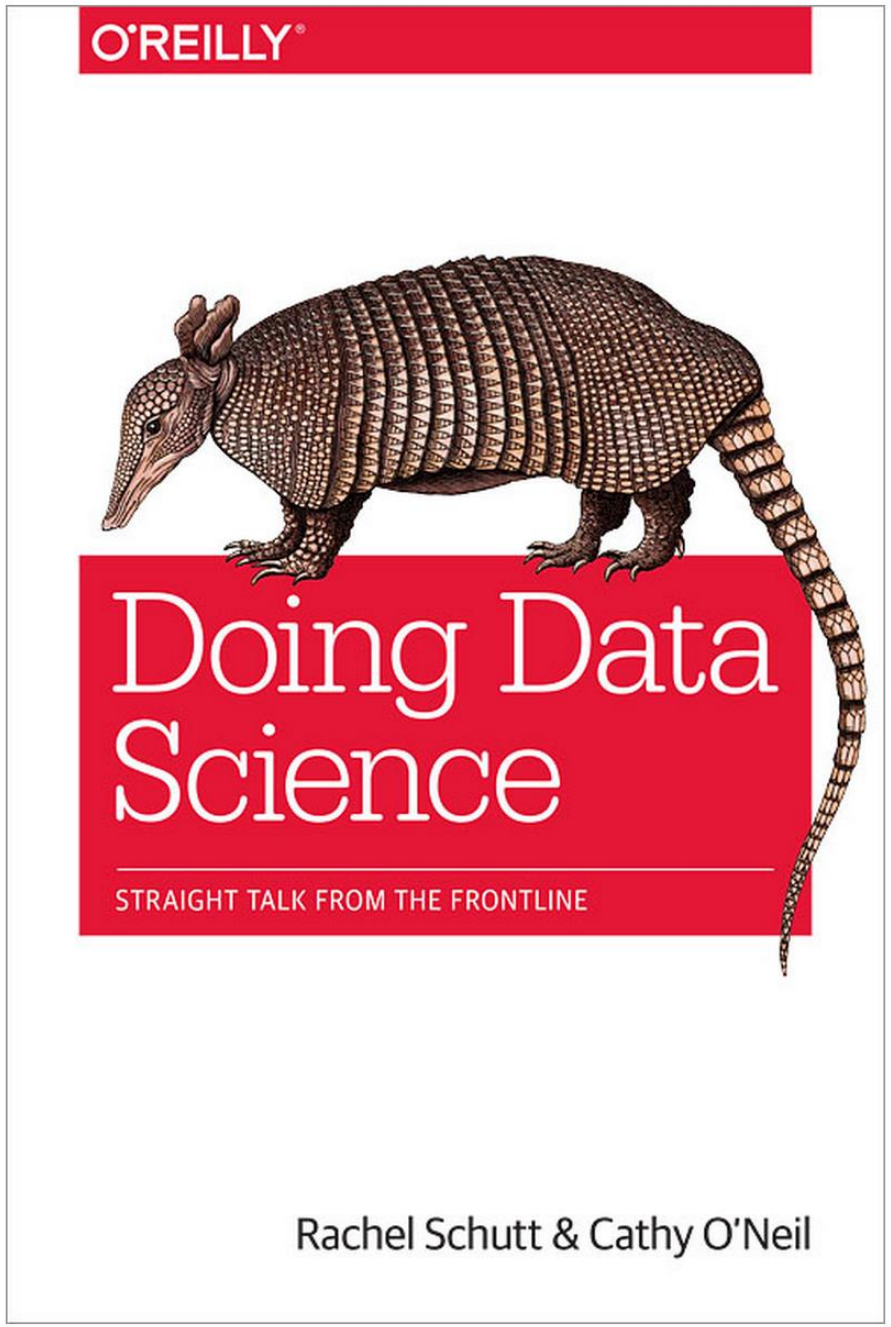


http://blog.infodiagram.com/2014/04/visualizing-big-data-concepts-strong.html

Big Data

Data science

Statistics



Four roles of a Big Data team

each provide different skills

RESEARCHER

Team member with experience applying formal research methods, including survey methodology and statistics

DOMAIN EXPERT

User, analyst, or leaders with deep subject matter expertise related to the data, its appropriate use, and its limitations

COMPUTER SCIENTIST

Technically skilled team member with education in computer programming and data processing technology

SYS ADMIN

Team member responsible for defining and maintaining a computation infrastructure that enalbes large scale computation

Tech roles

Computer Scientist

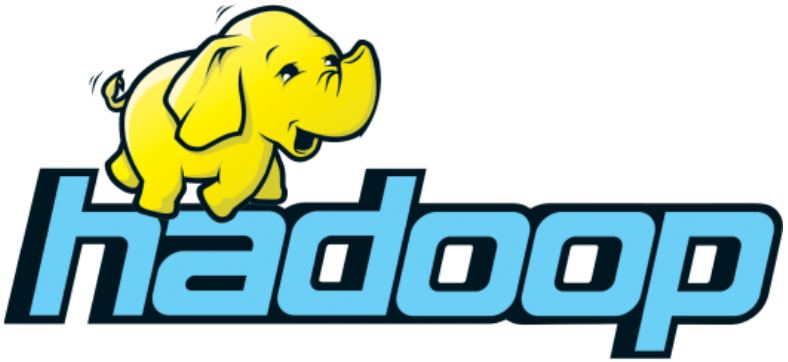
Distributed File Systems (DFS)

A 'doubly secure' storage solution

System Administrator

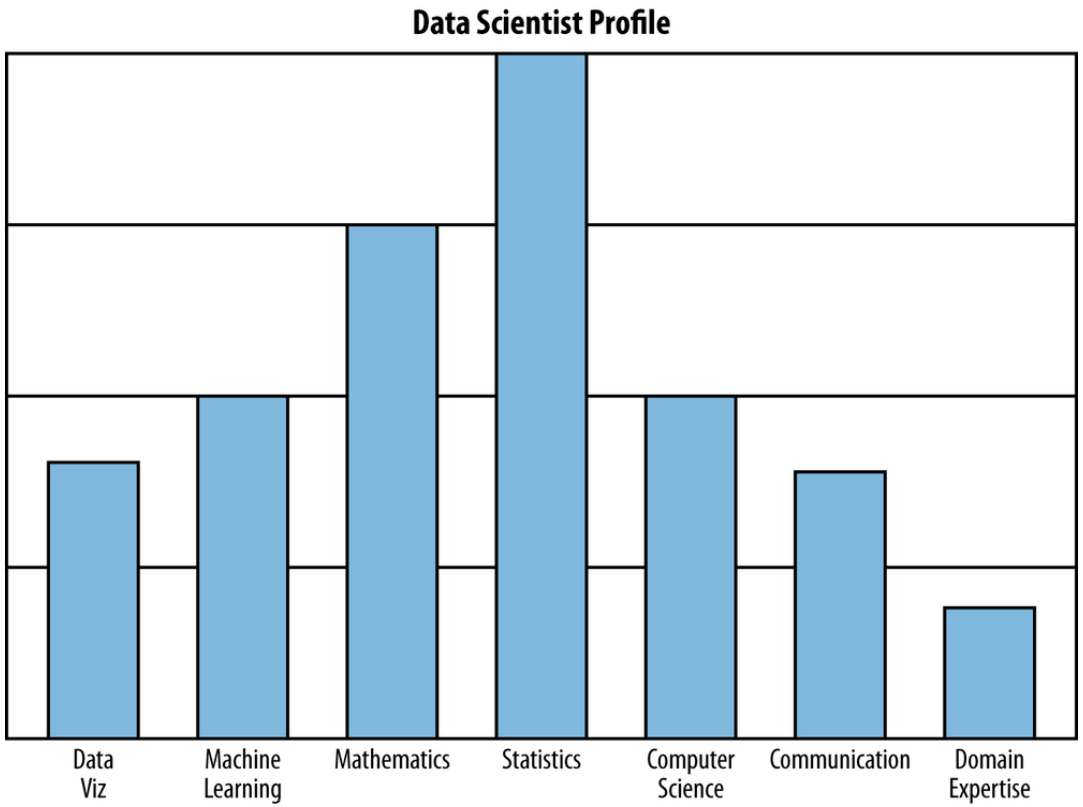
MapReduce (MR)

A 'divide and conquer' data processing model



Non-tech roles

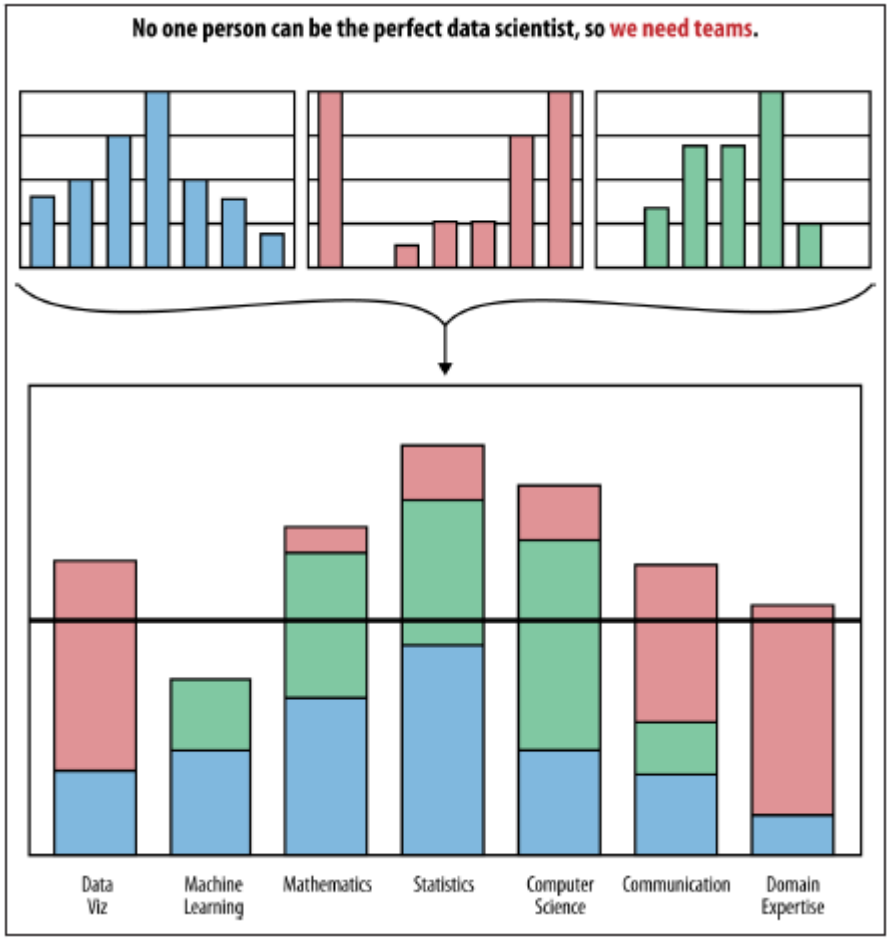
Domain Expert



Researcher

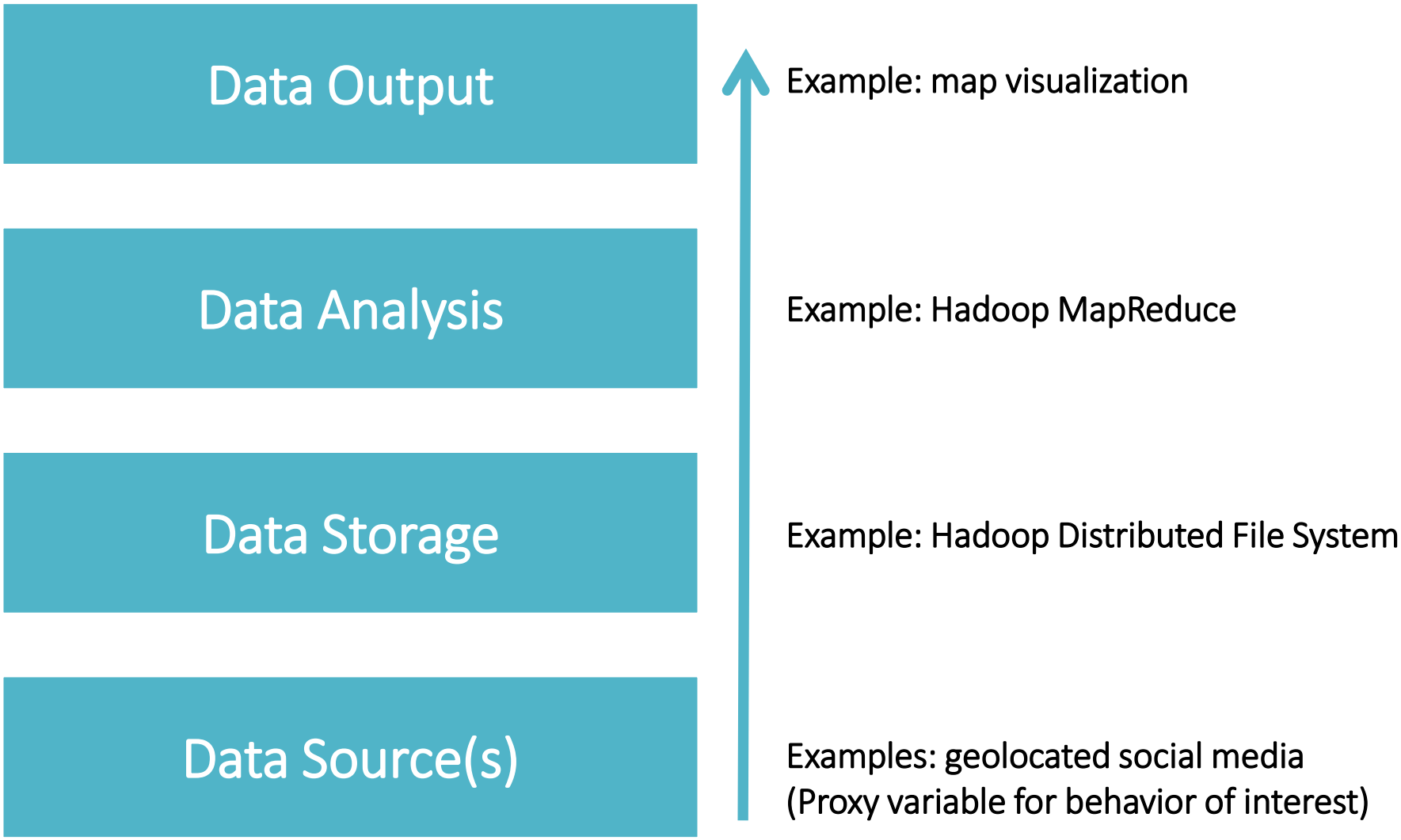
Source: Doing Data Science

The data scientist team



Source: Doing Data Science

Big Data: Layers



Source: Abe Usher

Thank you!

malex.berg@gmail.com