Introductions

DATA

Azza Abouzied
azza@nyu.edu
azzadev.github.io/databook

How much data (bytes) did we produce in 2010?

2 ZB

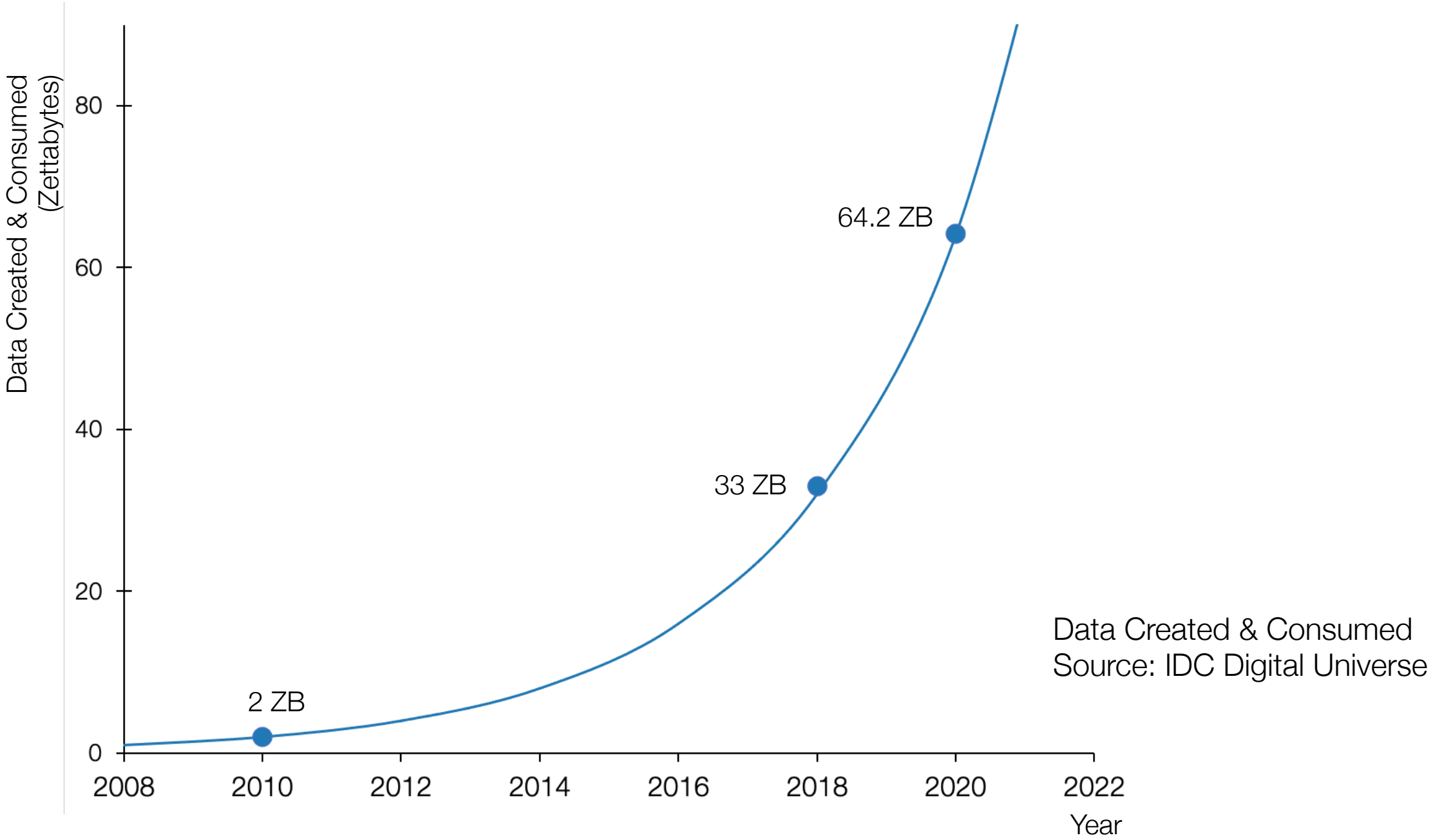
1 ZB = 1000^7 bytes = 10^{21} bytes = 1000 exabytes = 1 million petabytes = 1 billion terabytes = 1 trillion gigabytes.

How much data (bytes) did we produce in 2020?

64.2 ZB

In 10 years, our data grew by a factor of 32!

Our data roughly doubles every year!



Data Created & Consumed
Source: IDC Digital Universe

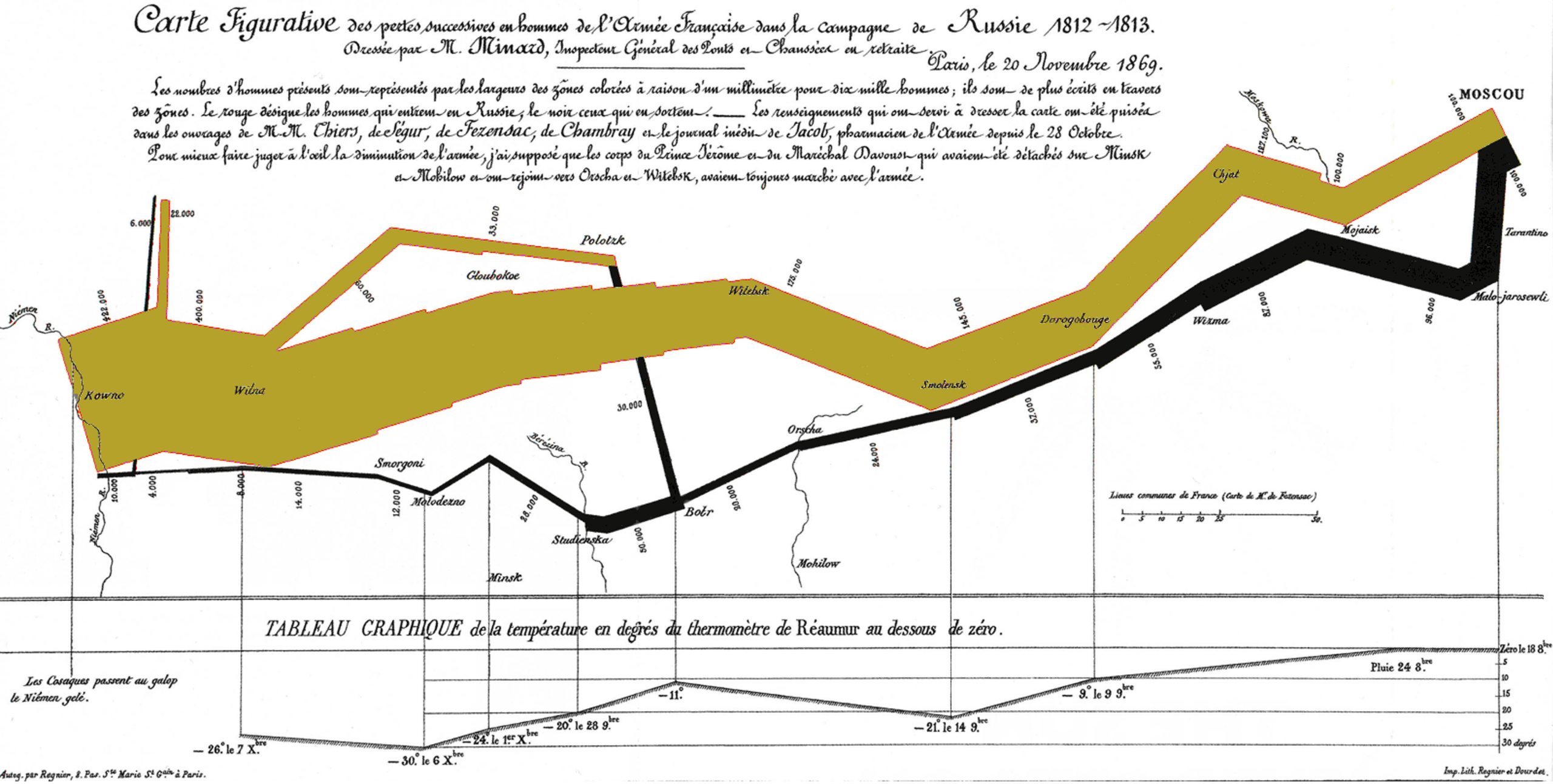
Where does this data come from?

Our interactions

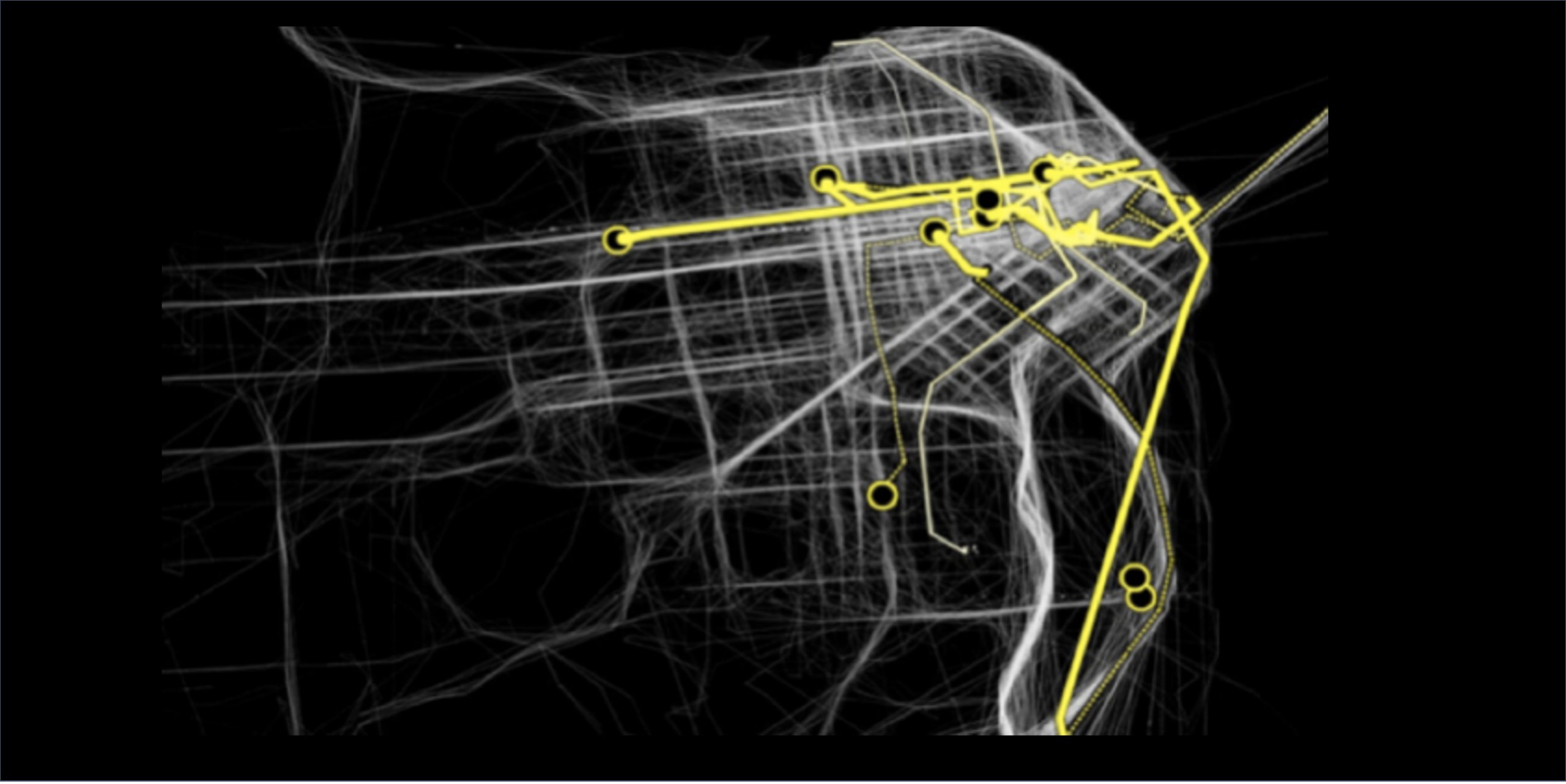
Our transactions

Heart Rate & Steps
21 SEP – 22 SEP
7:17a – 12:01a
9:00p
12:00a
3:00
7:17a
9:00a
12:00p
3:00p
6:00p
9:00p
12:01a
HEART RATE
Beats / Minute
High: 150
Avg: 69
Low: 30
STEPS
CALORIES
SKIN TEMP.
PERSPIRATION
180
160
140
120
100
80
60
40
20
0
8:00a 9:00a 10:00a 11:00a 12:00p 1:00p 2:00p 3:00p 4:00p 5:00p 6:00p 7:00p 8:00p 9:00p 10:00p 11:00p 12:00a
HEART RATE
STEPS
Steps / Minute
High: 115
Avg: 24
Low: 0
Total: 23551
CALORIES
SKIN TEMP.
PERSPIRATION
Sleeping
Awake
Active
Unknown

Our health


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 Mohilow en sont rejoints vers Orscha et Witebsk, avaient toujours marché avec l'armée.
Niémen R.
Kowno
Wilna
Smorgoni
Molodexno
Minsk
Studienska
Bobr
Berezina R.
Orscha
Mohilow
Polotzk
Gloubokoe
Witebsk
Smolensk
Dorogobouge
Wixma
Mojaisk
Chjat
Moskowa R.
MOSCOU
Tarantino
Malo-jarosewli
Lieues communes de France (Carte de M. de Fezensac)
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.
Les Cosaques passent au galop le Niémen gelé.
— 26.° le 7 X.bre
— 30.° le 6 X.bre
— 24.° le 1er X.bre
— 20.° le 28 9.bre
— 11.°
— 21.° le 14 9.bre
— 9.° le 9 9.bre
Pluie 24 8.bre
Zéro le 18 8.bre
Autog. par Regnier, 8. Pas. St.e Marie St. Gain. à Paris.
Imp. Lith. Regnier et Dourdet.

Our past



Our environment

Why take this course?

"The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades, ... because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it."

Hal Varian, Google's Chief Economist
The McKinsey Quarterly, Jan 2009

"Nothing that you will learn in the course of your studies will be of the slightest possible use to you in after life, save only this, that if you work hard and intelligently you should be able to detect when a man is talking rot, and that, in my view, is the main, if not the sole, purpose of education."

John Alexander Smith, Philosopher,
1914 Speech to Oxford University students.

Sounds familiar?

Data Power

Data Hubris

CHRIS ANDERSON SCIENCE 06.23.2008 12:00 PM

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

Illustration: Marian Bantjes “All models are wrong, but some are useful.” So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies [...]

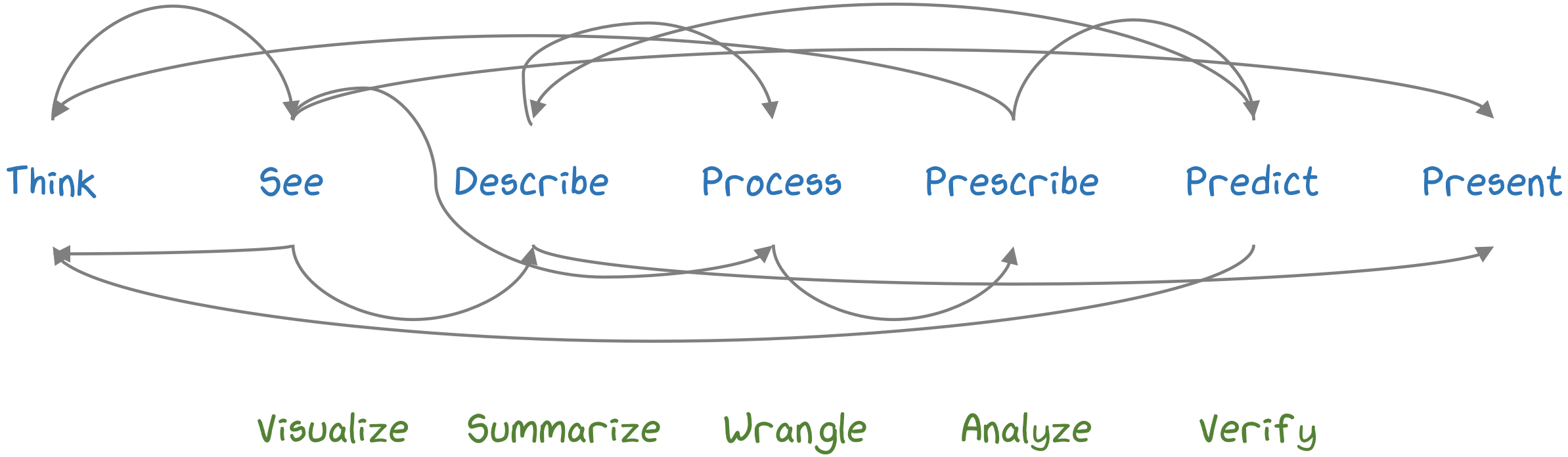
What is the road we will take?

NOT

A "Data Science" course

Find Download Clean Analyze End World Hunger Present

A "Critical Thinking" course



What I hope you learn?

Basics

- Data-driven thinking
- Limitations

Toolkit

- Excel / Spreadsheet software
- Tableau
- Trifacta
- Some IBM tools

How will I learn?

By questioning

By writing

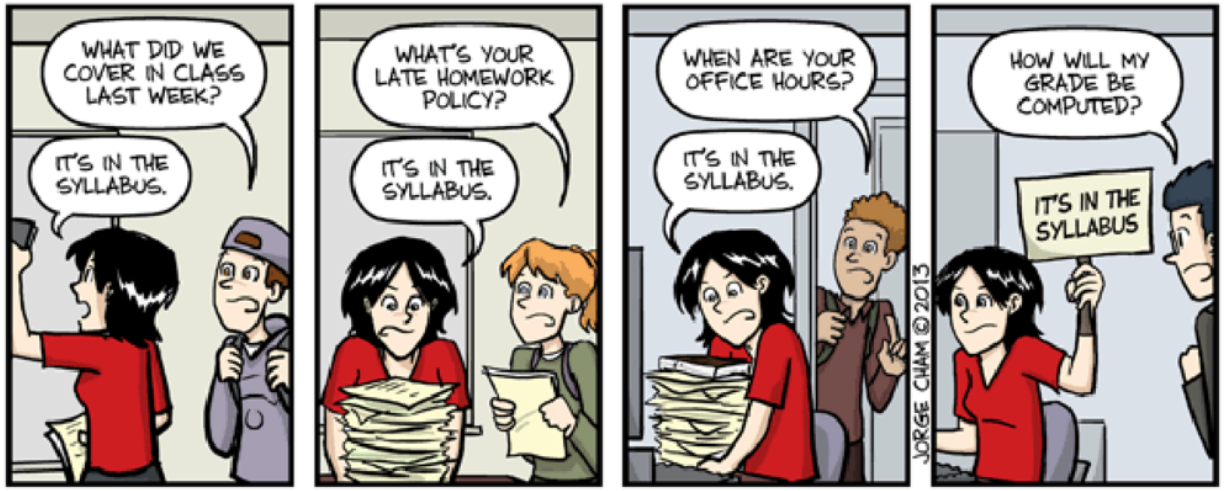
By doing

By collaborating

By peer feedback

By reading

By listening



IT'S IN THE ~~SYLLABUS~~

This message brought to you by every instructor that ever lived.

Syllabus & Logistics

COURSE BOOK

AZZADEV.GITHUB.IO/DATABOOK

Participation (the 5%)

Attendance is not enough

Engaging in class discussions

Engaging in outside class discussions

- Read a course reading, book, article, blog post, …
- How did influence it you?
- Write it a up in a post – doesn't have to be large (one reading and one post is not enough!)
- Share it with your peers
- Respond to your peers

Your very first exercise!
The Data Journal

Project Prep (for Wednesday)

Interests – what do you care about? What would you like to study deeply?

Knowledge – how much do you know about this domain? How much would you need to learn?

Skill set – what are your areas of strength and weakness?

Question – Propose an investigative/journalistic data project

Questions