

BIG DATA

A Revolution That Will Transform How We Live, Work and Think
Written By Viktor Mayer-Schonberger & Kenneth Cukier

Chapter One: Now

(Read This Introductory Chapter Pages 1-18 Weeks Two/Three in Class)

Google Predicts Flu

Airfare Forecasting

Telescopes (More Data, More Predictions)

More Data Forgives Flaws In Collection (e.g. Bias & Measurement Error)

Causality Gives Way To Correlation

Datafication, It's Not Just About Computers

Chapter Two: More

(Read in Class Week Three/Four – Two Assignments – Link with Sampling Distributions)

Population Versus Samples From A Population (Big Data Perspective – Pages 19-26)

Sumo Match Fixing & More Examples (Pages 26-31)

Chapter Three: Messy

(Read in Class Week Eight/Nine – Two Assignments – Link with Mathematical Modeling)

Changing The Goals of Data Collection (Pages 32-39 – Link with Sampling Design)

Big Data = More Comprehensive & More Frequent

Classification Less Rigorous (e.g. Identifying IBM)

Temperatures In A Field

Computer Chess Algorithms (Same Formulas But Better With More Data)

Computer Language Translators (Again Data Trumps the Algorithm)

Chapter Three: Messy (continued)

Departing From Exactitude (Pages 39 to 49 - [Link with Senior Survey & PFT Data](#))

BP's Cherry Point Refinery (Sensors)

Billion Prices Project (CPI Predicted in Real Time)

Tagging Photos on Twitter

Social Media "Likes" Counted/Reported (Less Classification, More Data)

Hadoop: Goodbye ETL (Leave It Where It Lies Vs. "Extract, Transfer, Load")

Zest Finance

Chapter Four: Correlation

(Read In Class Week Seven/Eight – Two Assignments – [Link with Bivariate Data Analysis](#))

All Hail Correlation (Pages 50 to 61 – [Link with Correlation - SMG](#))

Amazon's Marketing Strategy

Walmart's "Retail Link"

Blowing Off The Hypothesis ("Let The Data Speak" to Identify Variables)

FICO Scores

USPS: Commercial Vehicle Maintenance

Infrastructure Maintenance

Medical Diagnostics

The End of Theory-Based Decision Making (Pages 61 to 72 – [Link with Correlation](#))

Non-linear Correlations Now Possible

Does Causality Exist? (Philosophy in a Stat Class, What's Up With That?)

Debunking Louis Pasteur's Cure For Rabies

Exploding Manholes (Prediction Trumps Explaining Why)



Chapter Five: Datafication

(Read in Class Week Seven – Four Assignments)

History of Datafication (Pages 73 to 83 – Link with Observational Study Design)

Matthew Maury, Pathfinder of the Seas (Datafication Before Computer Digitization)

Shigeomi Koshimizu, Studying People’s Posteriors (Butt Prints Prevent Car Theft)

Double Entry Bookkeeping, “A Landmark . . . in the Use of Data”

When Words Become Data (Pages 83 to 86 – Link with Categorical Variables)

Google’s “Bold Plan” (Going Beyond Digitization To Datafication)

Amazon’s Statistically Significant Words

When Location Becomes Data (Pages 86 to 91 – Link with Quantification)

Datafication of Location: From Eratosthenes to GPS

UPS (“How Prediction Gave Us Knowledge”)

When Your Behavior Becomes Data (Pages 91 to 97 – Link with Study Design)

Wireless Big Brother Is Watching (Just Ask Eric Snowden)

Facebook’s “Social Graph” (Collecting Data Is Their Real Business)

What Twitter Knows

IBM’s Touch Sensitive Floor Covering (Tracking Your Every Step)

Chapter Six: Value

(Read In Class Week Nine/Ten – Three Assignments)

Data Becomes An Asset (Pages 98 to 107 – Link To Study Design)

ReCaptcha (Distinguishing Man From Machine)

Data’s Value = Easy to Collect, Cheap to Store, Abundant Uses, Long Shelf Life

Charging Stations For Electric Cars

Icebergs Help Explain the Value of Data Bases (Primary vs. Secondary Uses)

Amazon Gets Over On AOL,

Google “1” Nuance “0” (Using the Information Value Chain)



Chapter Six: Value (Continued)

Recombinant Data (Combining Data Sets Leads To Value of Whole > Sum of the Parts)

Do Cell Phones Cause Cancer? (Recombinant Data Provides An Answer)

Expanding Extensibility (Flexible Data Collection Looks Beyond Primary Use)

Retailers Install “More Than Security” Cameras

Depreciating Value of Data

Spell Checking Software (How Google Did It Better Than Microsoft)

Public Sector Data Bases (Government’s Role in Data Collection & Distribution)

“Open Government Data” Initiatives (Just Let The Private Sector Do It)

FlyOnTime.us

Facebook’s IPO Value Vs. Book Value (“Option Value of Data”)