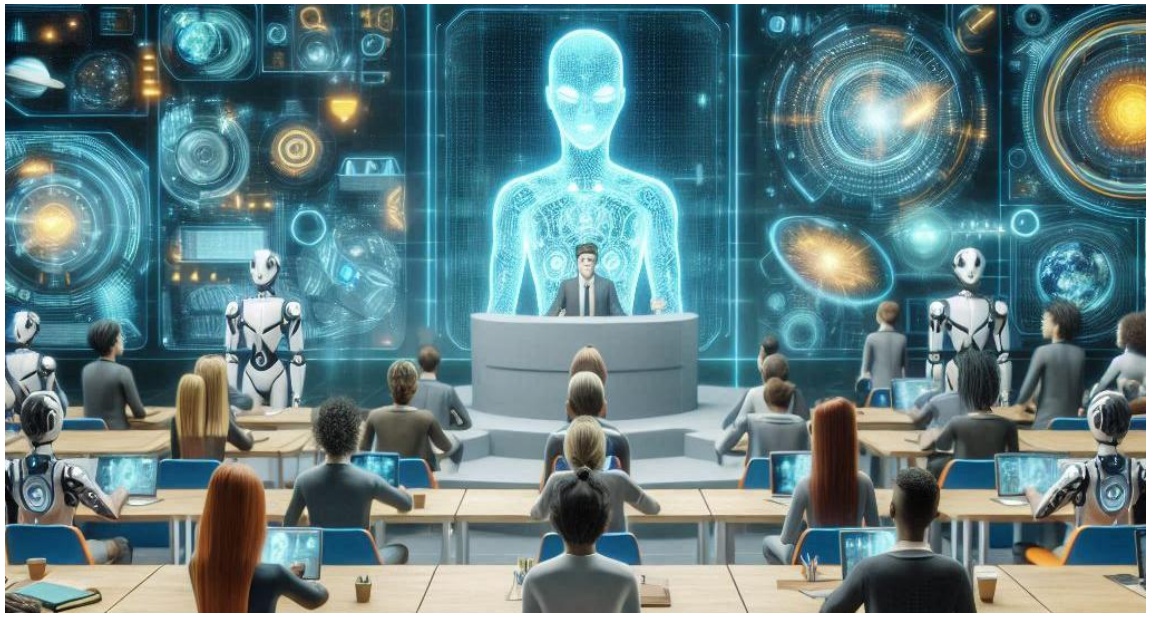
CS 180 INTRODUCTION TO DATA SCIENCE

INTRODUCTION, COURSE OVERVIEW, OBJECTIVES, HOW TO SUCCEED



Created by DALL-E Prompt: “Artificial Intelligence Classroom”

SOME ANNOUNCEMENTS

OPEN HOUSE for 5x AI & DATA SCIENCE MAJORS!

Pop in to grab food, ask questions,
meet the new Data Science Administrator,
and learn about new majors, clubs, and events!

Sept 3rd-5th, 8th-10th
8am-5pm
West View Building (WVB) #1162

STEM career fair

PREP WORKSHOP

Thurs., September 11
12:00 PM – 1:00 PM
EB 204

FAIR!

Thurs., September 18
9:00 AM – 3:00 PM
WSC Ballroom & Garden Court

Connect with top employers and secure your next job or internship!

BYU Career Services

BYU Statistics
BYU Career Services

Data Science and Actuarial Career Fair

Wed., October 1
9:00 AM – 12:00 PM
Lunch*: 12:00 PM – 1:30 PM
Hinckley Center Assembly Hall

Register here!

*RSVP required for lunch.

ANNOUNCEMENTS

Dr. Jake Rhodes

rhodes@stat.byu.edu

WVB 2177

Office Hours: TBD

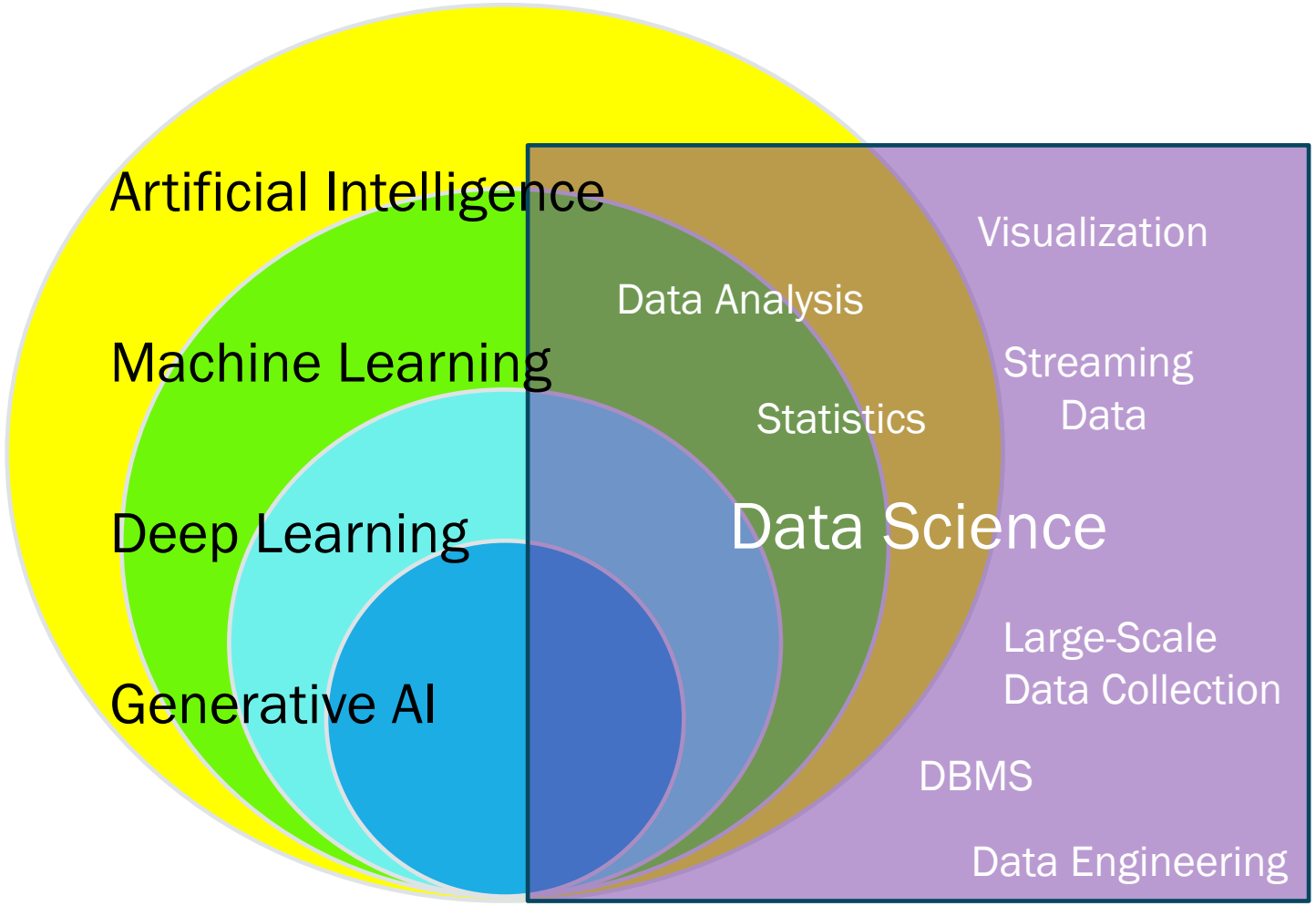
Teaching Assistants (WVB 1151):

1. Toby Allen
2. Corbin Christiansen
3. Michael Jensen
4. Spencer Marshall
5. Ian Villanueva

Office Hours: Posted on Syllabus

BYU

WHAT IS DATA SCIENCE?



Data Science:

- **Scope**: focuses on data as a whole, including data collection, processing, analysis, storage, and management.
- **End Goals**: Primarily concerned with extracting knowledge and actionable insights from data.
- **Techniques**: Uses data collection, data cleaning, data transformation, statistical analysis, data visualization, data management and data engineering tools.

BYU

WHY DATA SCIENCE?


Healthcare


Advertising

Finance


Genomics


Augmented reality


Education


Speech recognition


Data cleansing


Financial fraud detection


Logistics


Video game


Retail

Data science in advertising


Manufacturing

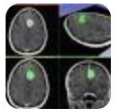

Transportation


Sports


Internet search


E-commerce


Airline route planning


Medical image computing


Drug development


Recommendation systems

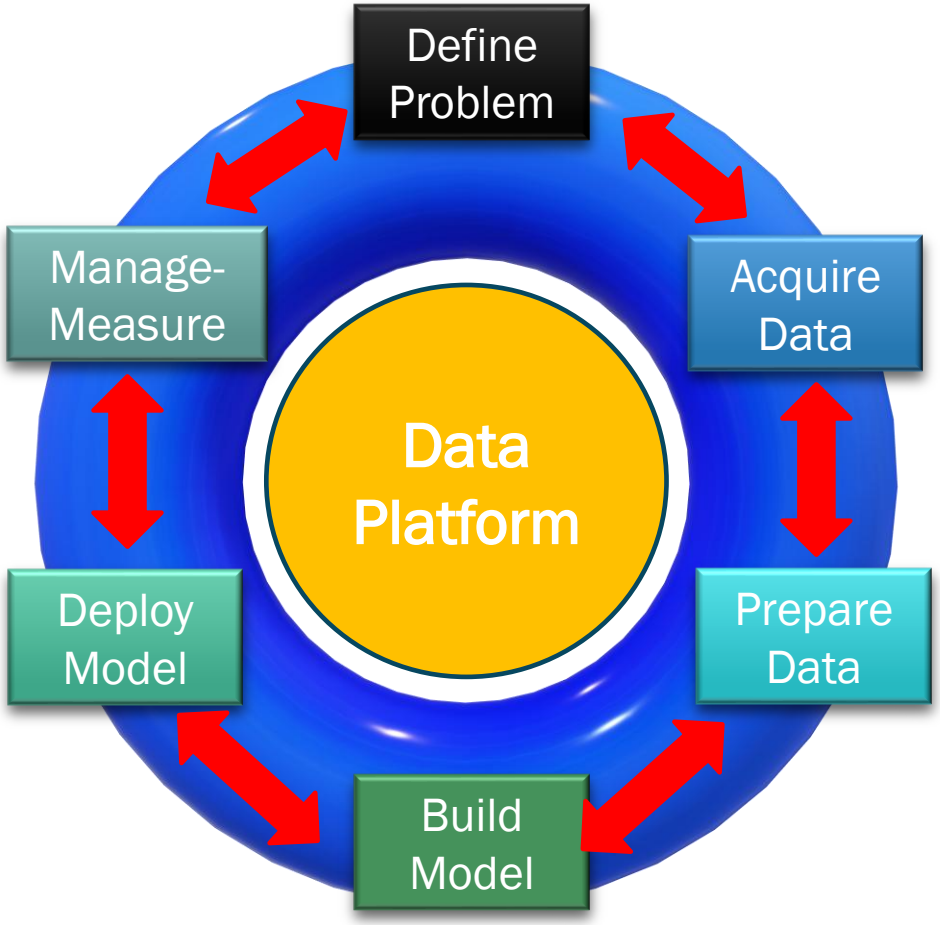

Energy


Virtual assistants

BYU

TECH STUFF WE'LL LEARN

- The Data Science Lifecycle process
- Using Python for data science (Numpy, Pandas, Matplotlib, Scikit-Learn, and more)
- How to prepare data for analysis
- How to explore data for insights
- Data Visualization (Python and Tableau)
- Data Management (SQL)
- Machine Learning basic algorithms
- Use of GenAI tools for language-based problems



BYU

GEN AI APPLICATIONS

Code

Video

Text

Image

Audio/Music

suno.com

BYU

WE WILL COVER A LOT THIS SEMESTER!

Class #	Week #	Month	Date	Topic	Reading	Labs	Due date
1	1	Sep	4	Welcome, Introduction, Course Objectives, DS Lifecycle	Chapter 1 Intro DS	Lab 1: IDE Set Up, GitHub	Sept 6, 2025
2	2	Sep	9	Python setup, Google colab, Github	Chapter 2.1-2.5 Python		
3	2	Sep	11	Numpy, vectorization, reading/writing		Lab 2: Vectorization	Sept 13, 2025
4	3	Sep	16	Pandas, Matplotlib, Seaborn	Chapter 2.6-2.8 Python		
5	3	Sep	18	Data Cleaning and Preparation		Lab 3: Pandas, Numpy	Sept 20, 2025
6	4	Sep	23	Data Acquisition, ETL, Populations, Sampling	Chapter 3 Data Prep		
7	4	Sep	25	Descriptive Statistics	Chapter 4 Prob & Stat	Lab 4: Data Preparation	Sept 27, 2025
8	5	Sep	30	Exploratory Data Analysis (EDA)	Chapter 5 EDA		
9	5	Oct	2	Principles of Data Visualization		Lab 5: Data Visualization	Oct 04, 2025
10	6	Oct	7	Business Intelligence with Tableau	Chapter 6 BI & Tableau		
11	6	Oct	9	Tableau Dashboards, Stories		Lab 6: Data Story Telling with Tableau	Oct 11, 2025
12	7	Oct	14	Data management - databases, SQL queries			
13	7	Oct	16	More SQL Features, Joins	Chapter 7 DB and SQL		
14	8	Oct	21	SQL continued, NoSQL, Connect to Tableau			
15	8	Oct	23	MIDTERM REVIEW		Midterm	Oct 25, 2025
16	9	Oct	28	Overview of AI, ML, DL, GenAI Topics for remainder of semester	Chapter 8 Unsupervised Learning		
17	9	Oct	30	Unsupervised Learning- Kmeans		Lab 7: Data Engineering with SQL	Nov 1, 2025
18	10	Nov	4	Unsupervised Learning- Hierarchical, DBSCAN	Chapter 9 Supervised Learn		
19	10	Nov	6	Supervised Learning: Part 1	Chapter 10 Decision Trees	Lab 8: Cluster Analysis	Nov 08, 2025
20	11	Nov	11	Supervised Learning: Part 2	Chapter 11 Regression (optional)		
21	11	Nov	13	Evaluation of models, comparing performance	Chapter 12 Eval	Lab 9: ML Classification/Regression	Nov 15, 2025
22	12	Nov	18	ANN, Multi-Layer Perceptron, Backpropagation	Chapter 13 ANN		
23	12	Nov	20	Deep Learning		Lab 10: MLP and Backpropagation	Nov 22, 2025
24	13	Nov	25	GenAI - Introduction	Chapter 14 GenAI		
-	13	Nov	27	No Class. Thanksgiving Holiday.			
25	14	Dec	2	GenAI - Applications	Chapter 15 AI Ethics		
26	14	Dec	4	Ethics / Data Privacy / Business and Government Policy		Lab 11: GenAI Applications	December 06, 2(
27	15	Dec	9	Review and Wrap Up, The Future of DS/ML/AI			
28	16	Dec	17	Finals Week. Final 12/12 2:30-5:30pm 1102 JKB		Final	

BYU

COURSE GOALS

- Enhance your ability to quickly learn and master new concepts
- Utilize cutting-edge data science languages and tools to extract insights from data
- Effectively communicate insights through data visualization, dashboards, data storytelling
- Think critically about conclusions drawn from data analysis
- Recognize the potential to do good by applying Data Science to address critical global problems

BYU

WE ARE A LEARNING COMMUNITY

- Definition: A group of people who come together with the shared goal of learning, growing, and improving—often through collaboration, discussion, and mutual support. [ChatGPT]
- Get to know each other. How?
- At the beginning of every class, introduce yourself to those around you.
- Benefits: More effective learning, better performance, build a social network, better sense of belonging, and it’s more fun!

BYU

AI “POLICY”

- Generative AI tools like ChatGPT, Gemini, Claude, Grok, etc., can be helpful to assistants in the learning process.
- If misused, GenAI can become a crutch, lead to sloppy thinking, and even a substitute for engagement with real people.
- We want you to have the best learning experience, resulting in foundational understanding and deep skill development.
- We recommend limiting GenAI use to answering questions about topics, looking up syntax for code, and understanding code examples, but not using GenAI to answer coding questions end-to-end.
- We trust you will follow this recommendation (aka policy). We will not police you, but if it is obvious, we will bring this to your attention.
- To help you better learn, we will have portions of the mid-term and final that you can’t do with AI. As a result, it will benefit you to know the concepts to the point where you do not need AI.

BYU

UPCOMING ASSIGNMENTS

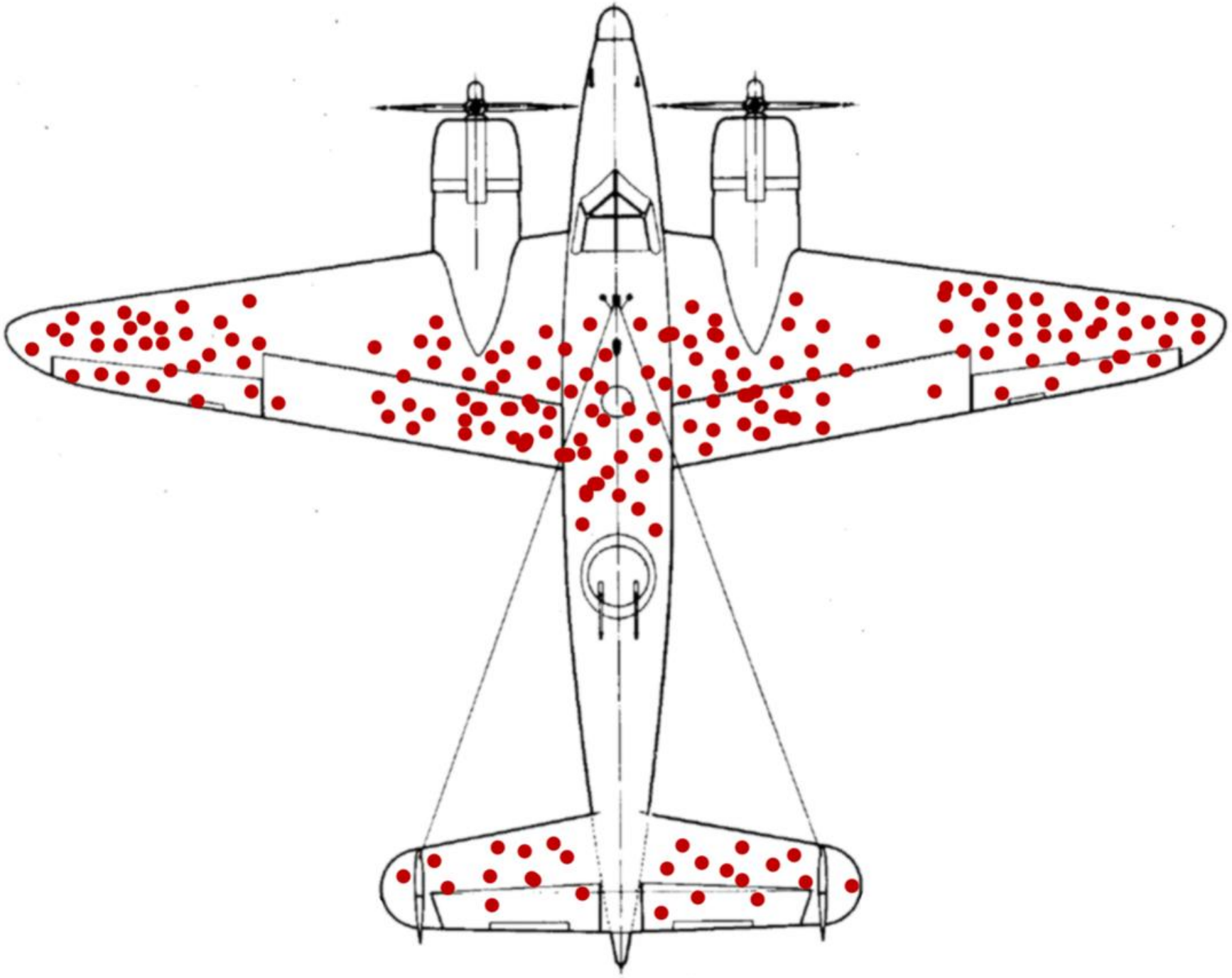
1. Sign up for zyBook. How? Go to the first reading assignment in Canvas and click “Load Chapter 1 Reading…” This will initiate the process of buying the zyBook. You should not have to do this again.
2. Reading Assignment: Chapter 1. Points automatically accrue as you do the activities in the book.
3. Data Science Lab 1: Intro to Colab. Set up a Python Development Environment. Programming assignments will be turned in via Google Colab notebooks. However, if you are planning a career requiring programming, I recommend using a professional IDE, such as VS Code. If needed, we’ll go through the setup quickly next week.
4. Getting Set Up “Assignment”: This will help you (and me) hit the ground running!

BYU


+ableau
102 VISUZLIATIONS
VIZ LIBRARY
THE COLLECTION OF TABLEAU VISUALIZATIONS & ANALYSIS
V.2018.8.3
Kritidikoon Woraitthinan
KPI
STATIC (3)
COMPARISON (8)
TREND (14)
CONTRIBUTION
STATIC(3)
COMPARE(4)
TREND(3)
RANKING
STATIC (3)
COMPARISON (3)
TREND (2)
AMOUNT (5)
FLOW (9)
TIME (7)
RELATIONSHIP (10)
DISTRIBUTION (6)
GROUPING (5)
TEXT (2)
MAP (8)
EXTRA (2)
★1=SIMPLE ★2=MODERATE ★3=DIFFICULT ★4=ADVANCED ★5=MASTER
How-to-Build / Credits
KPI VALUE - STATIC【KPI値 - 単一】
Gauge1 ★2
Gauge2 ★3
Number ★1
95%
475
95%
+75
KPI VALUE - COMPARE【KPI値-比較】
Radial Gauge ★3
Circular Radial ★4
Radial Bar ★5
Likert Scale ★3
Scale ★2
Radar Chart ★3
Double Bar ★1
Bullet ★1
KPI VALUE - TREND【KPI値-推移】
Bar Chart ★1
Bar & Line ★2
Bar & Shape ★2
Bold Line ★1
Gap Trend ★3
Waterfall ★3
Lollipop ★2
Spiral ★3
Line Chart ★1
Line Cumu ★1
Align Start ★3
Sparklines ★1
Multiple Chart ★1
Forecast ★1
CONTRIBUTION - STATIC【貢献 - 単一】
Pie Chart ★1
Donut Chart ★2
Waffle ★2
CONTRIBUTION - COMPARE【貢献 - 比較】
Tree ★1
Bubble ★1
Growth Ring ★2
Contri. Bar ★1
Butterfly ★2
CONTRIBUTION - TREND【貢献 - 推移】
Stacked Bar ★1
Area ★1
FLOW【フロー】
Funnal1 ★2
Funnal2 ★3
Snakey ★4
Gantt Chart ★1
Process Flow ★1
Decison Tree ★4
Network ★4
Network2 ★4
RELATIONSHIP【関係性】
Heat Map1 ★1
Heat Map2 ★1
Pereto ★2
RANKING- COMPARE【ランキング-比較】
Ranking Board ★2
1 Kenji
2 Maki
3 Tomomi
Rank-Race ★2
Star Rating ★2
Arrow ★2
Slopegraph ★2
Slopegraph 2 ★3
RANKING- TREND【ランキング-推移】
Bump Chart1 ★2
Bump Chart2 ★2
TIME【時間】
Timeline ★3
Multi-timeline ★3
Calendar ★2
Clock ★2
Jump Plot ★4
Jump Plot2 ★4
Circular Dot ★3
DISTRIBUTION【分布】
Histrogram ★1
Box Plot ★1
Scatter Plot ★2
Scatter Plot2 ★2
Jitters ★1
Barcode Plot ★1
Correlation ★2
Scatter Plot ★1
Reg-Linear ★1
Reg-Log ★1
Reg-Expo. ★1
Polynomial ★1
GROUPING【グルーピング】
Venn Diagram ★2
Venn Diagram ★3
Sunburst ★4
Quardiant ★2
Culstering ★1
Bubble Plot ★1
TEXT ANALYTICS【テキスト分析】
Word Cloud ★2
Word Count ★2
Sentiment Analysis ★4
MAP【地図】
Map - Filled ★1
Map - Symbol ★1
Map - Pie ★2
Map-Custom ★3
Path - Route ★3
Path - Direct ★3
Path - Curve ★4
Path - Curve Multi-line ★4
Position ★3
EXTRA【特別】
3D Model ★5
Geometric Art - Drawing ★5
Shape- Heart ★3
Let’s Dive In!

- What do you see?
- What should you do?
- Analyze or act?

BYU

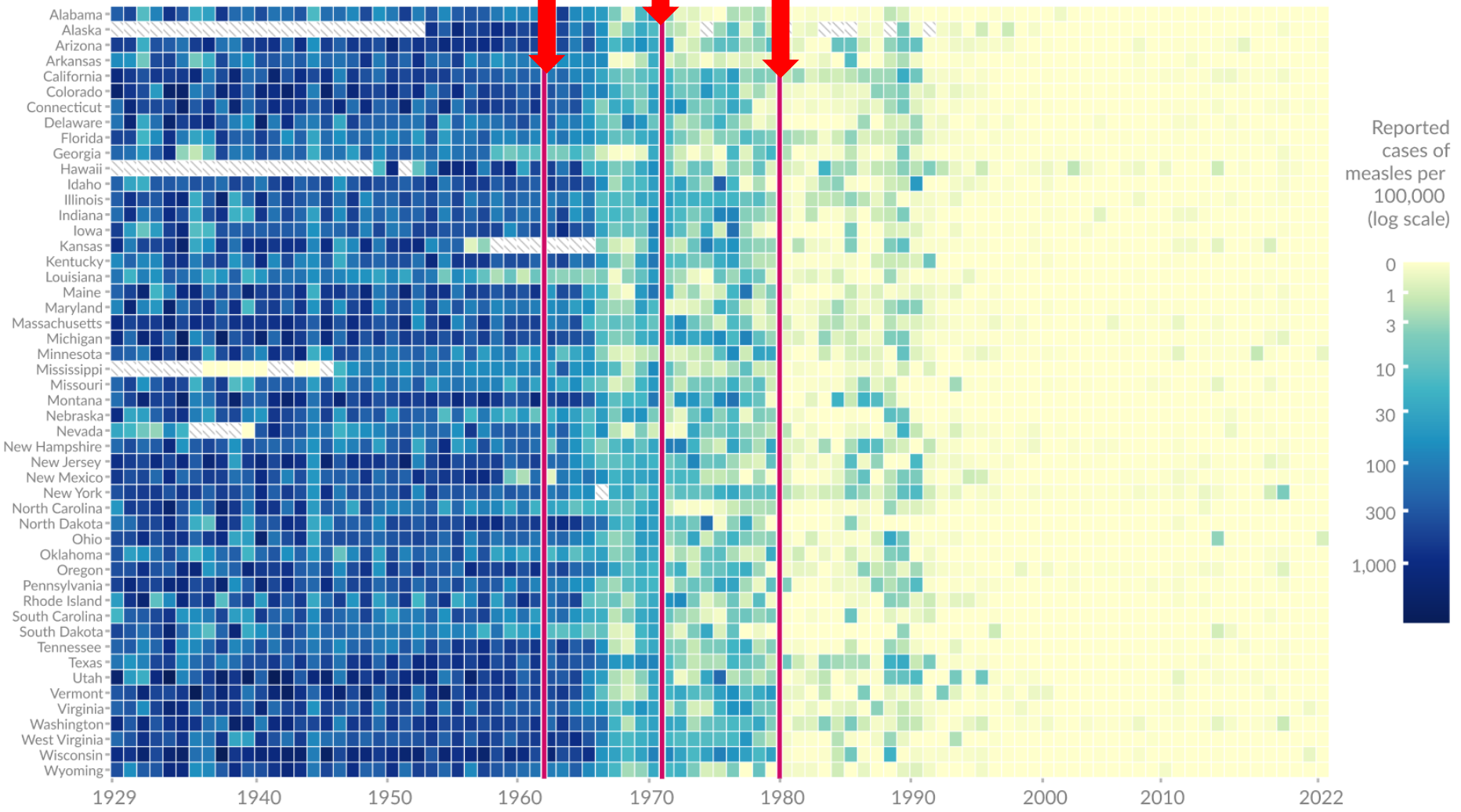


1971: Measles-mumps-rubella vaccine developed

1963: Measles vaccine developed

1980: Mandatory for kindergarteners to be vaccinated

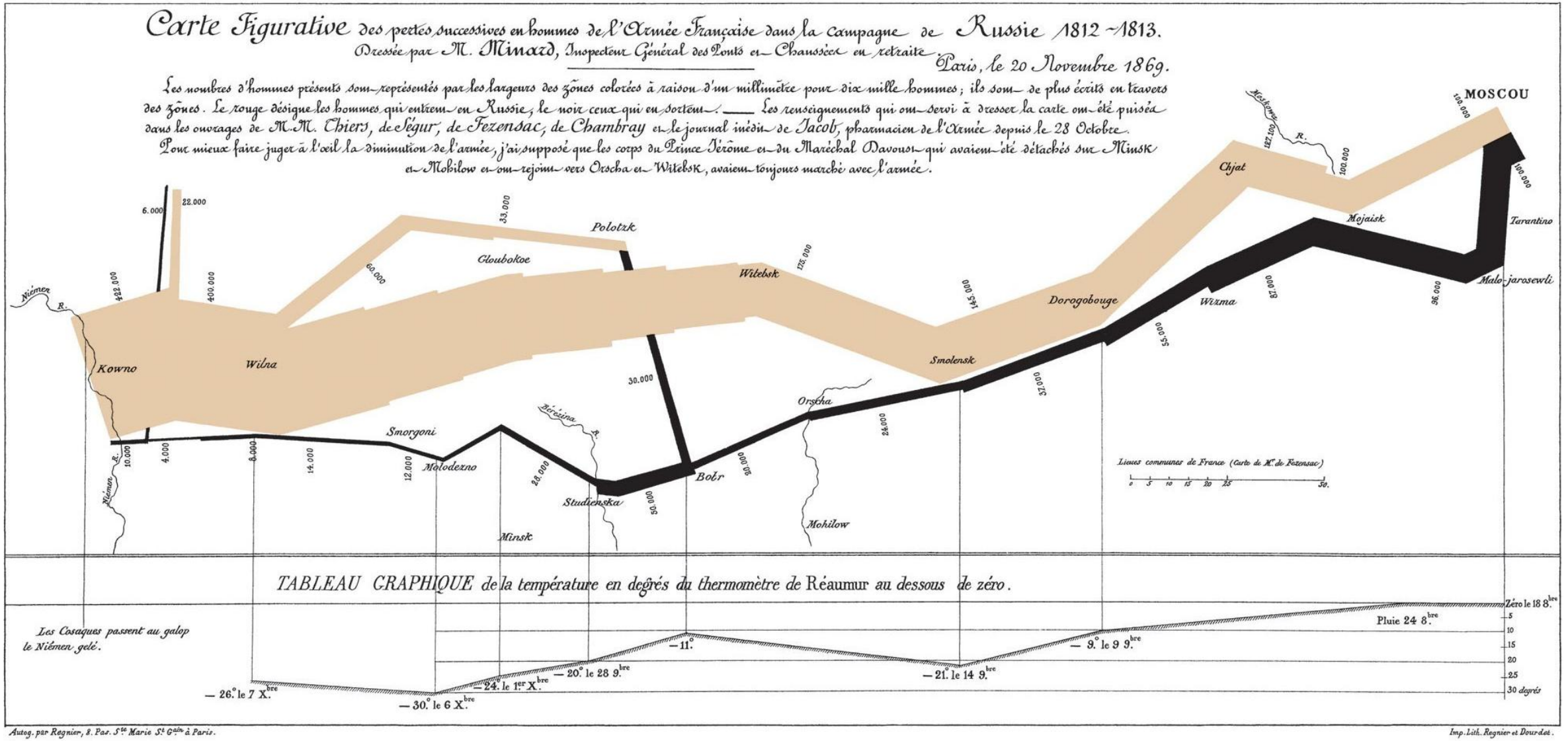
US States



Data source: Project Tycho (2018); Centers for Disease Control and Prevention (1959–2022)

BYU

NAPOLEON'S DISASTROUS INVASION OF RUSSIA IN 1812



GENDER BIAS AT BERKLEY (1973)

Are men applying to Berkeley more likely to get in than women?

	Men		Women	
	Applicants	Admitted	Applicants	Admitted
Total	8442	**44%**	4321	35%

BYU

GENDER BIAS AT BERKLEY (1973)

Are men applying to Berkeley more likely to get in than women?

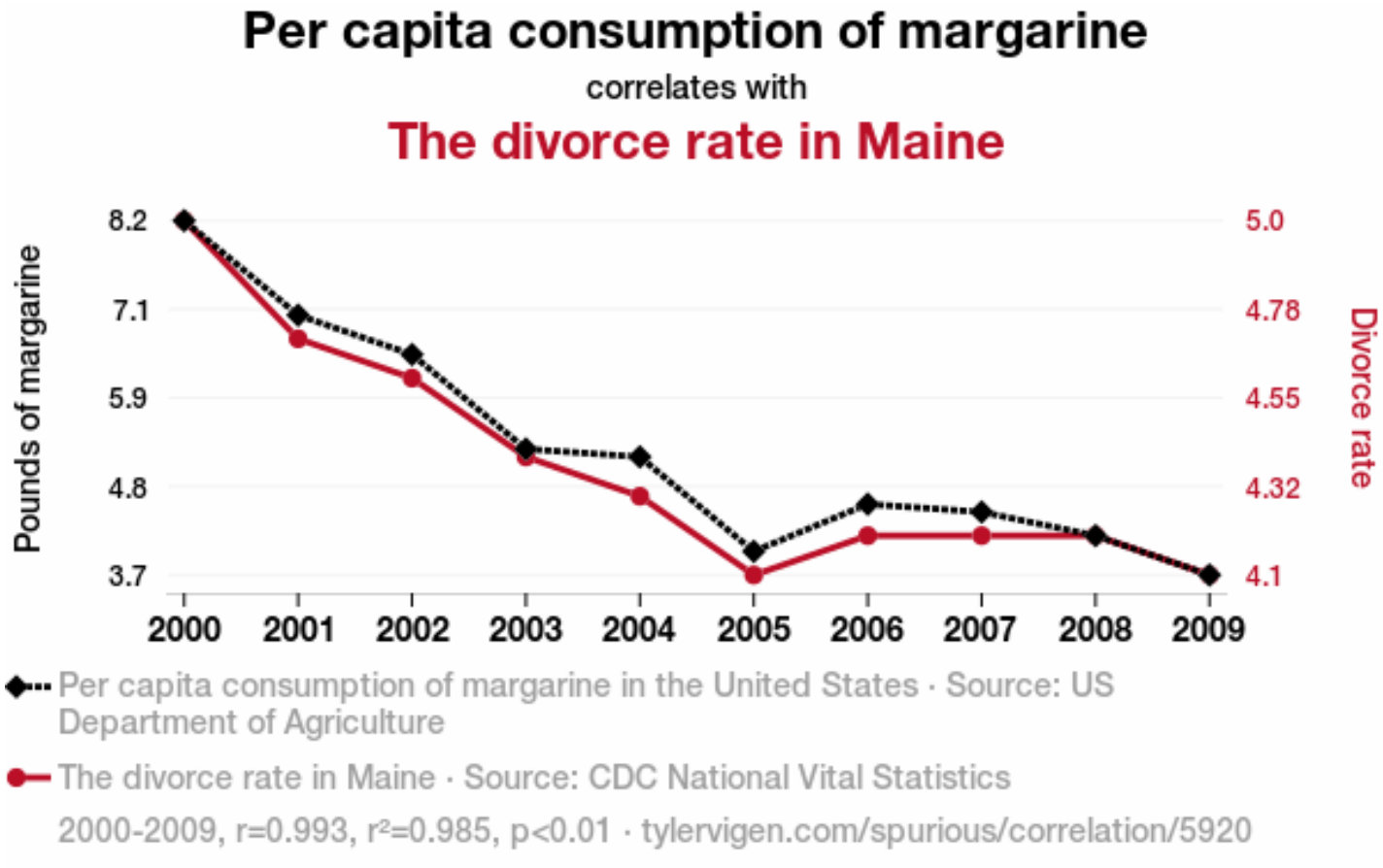
	Men		Women	
	Applicants	Admitted	Applicants	Admitted
Total	8442	**44%**	4321	**35%**

- *Departments have different acceptance rates,*
- *More women applied to departments with lower acceptance rates*

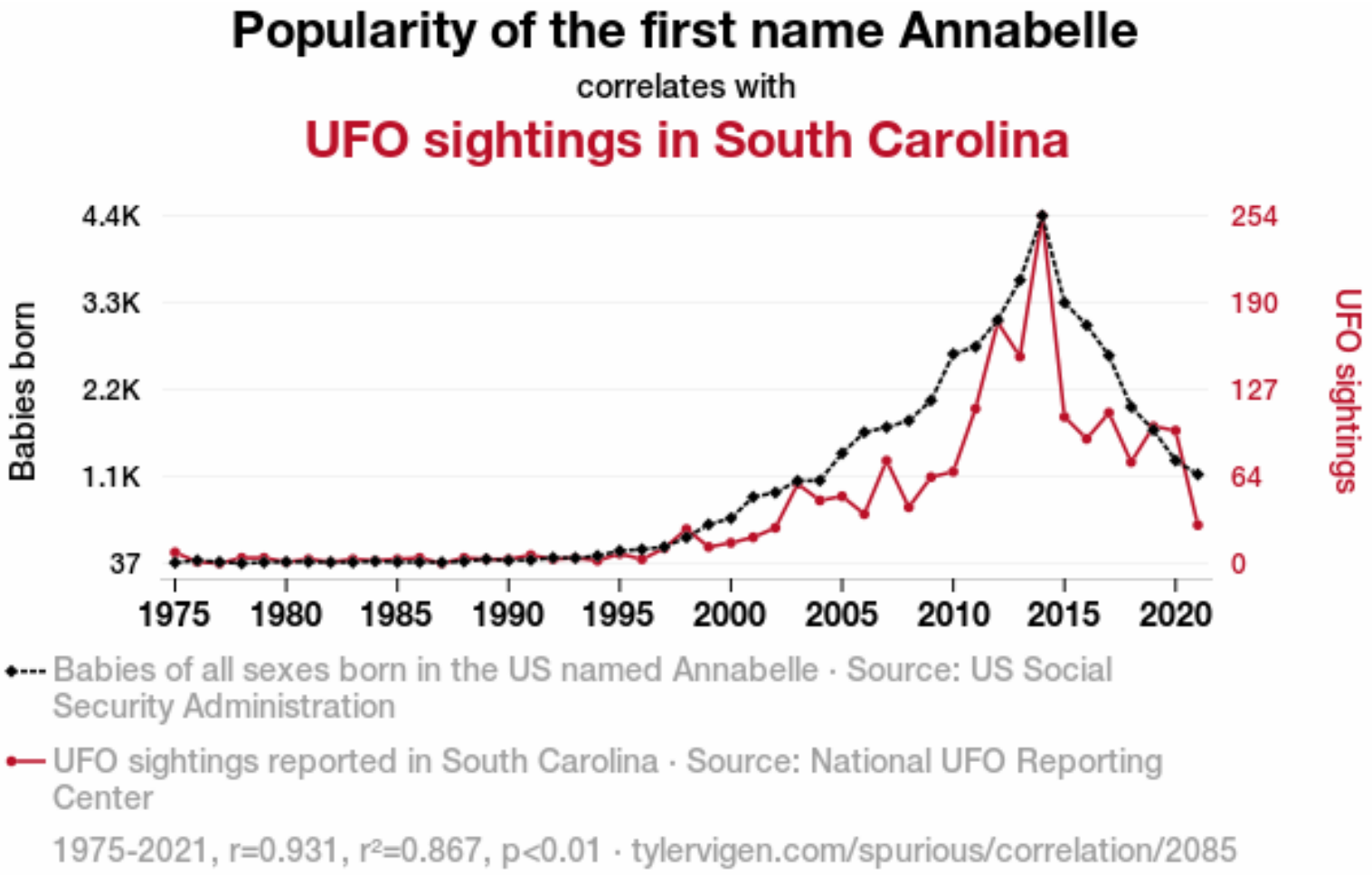
Department	Men		Women	
	Applicants	Admitted	Applicants	Admitted
A	***825***	62%	108	**82%**
B	***560***	63%	25	**68%**
C	325	**37%**	***593***	34%
D	417	33%	375	**35%**
E	191	**28%**	***393***	24%
F	373	6%	341	**7%**

SPURIOUS CORRELATIONS

SPURIOUS CORRELATIONS



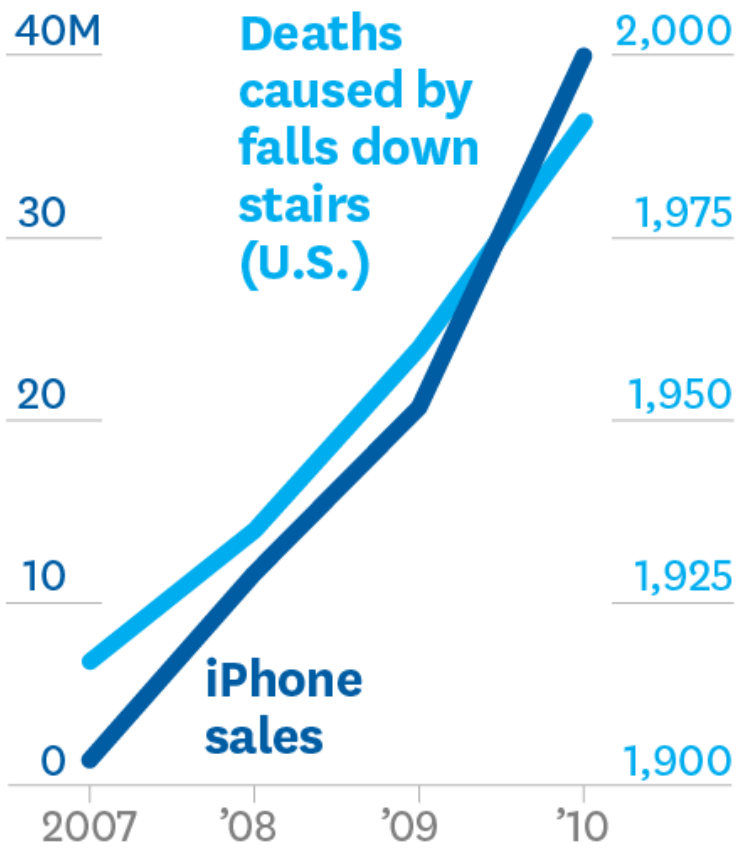
SPURIOUS CORRELATIONS



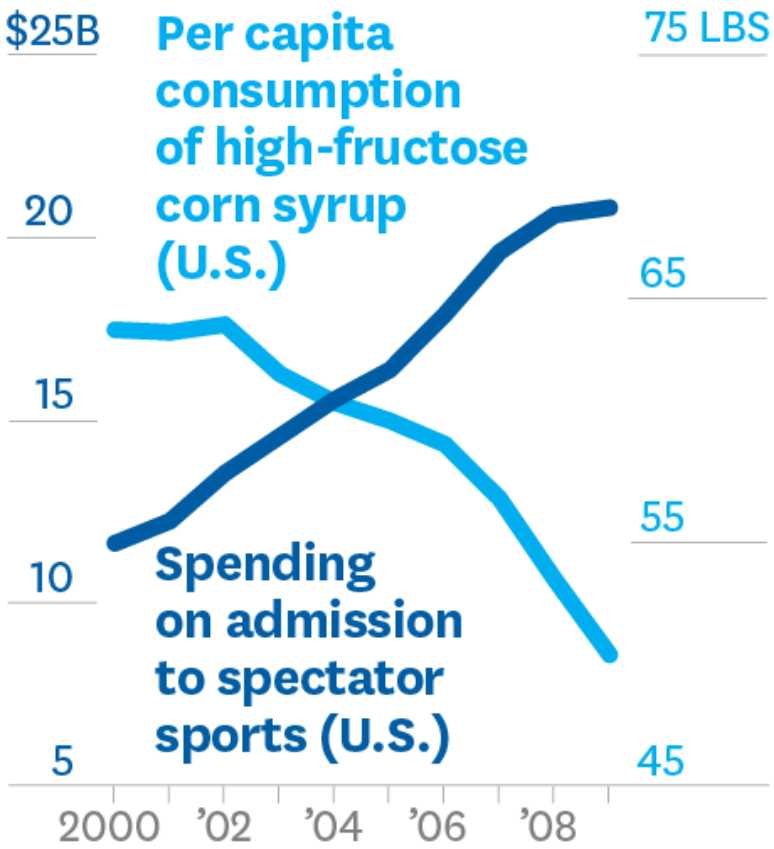
https://www.tylervigen.com/spurious-correlations

BYU

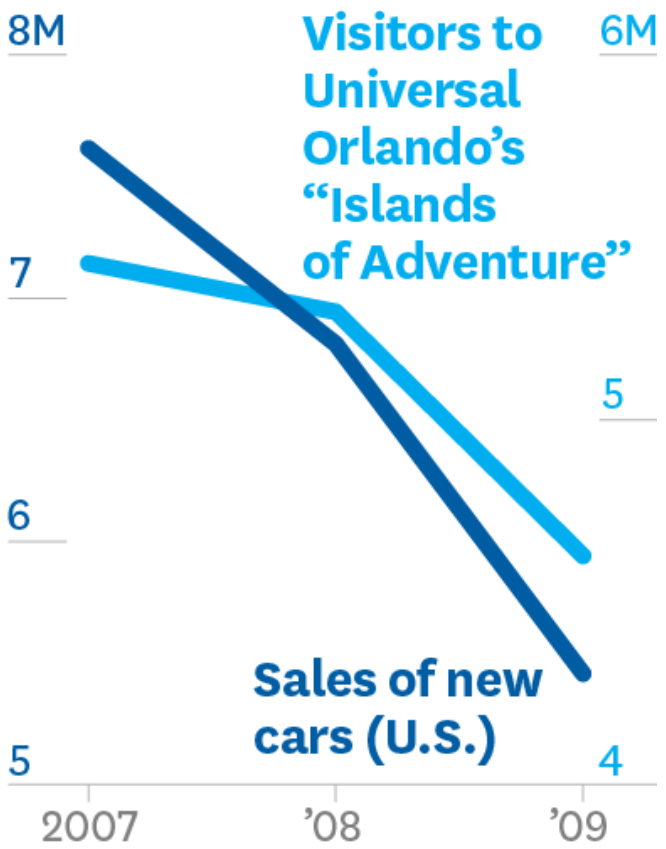
MORE IPHONES MEANS MORE PEOPLE DIE FROM FALLING DOWN STAIRS



LET'S CHEER ON THE TEAM, AND WE'LL LOSE WEIGHT



TO INCREASE AUTO SALES, MARKET TRIPS TO UNIVERSAL ORLANDO



SOURCE TYLERVIGEN.COM
FROM "BEWARE SPURIOUS CORRELATIONS," JUNE 2015

ZYBOOK DATA SCIENCE LIFECYCLE FOR DATA ANALYSIS

Table 1.4.1: Data science lifecycle.

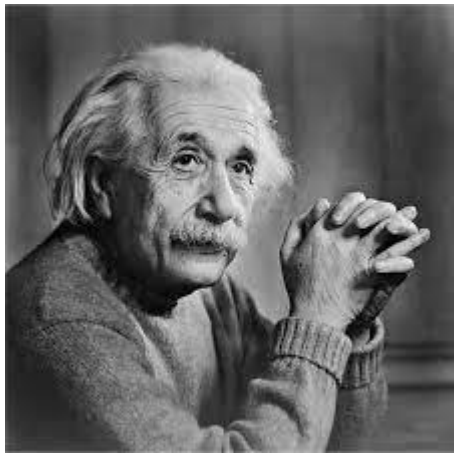
Step	Description
Step 1: Gathering data	Identify available and relevant data; gather new data if needed.
Step 2: Cleaning data	Reformat datasets, create new features, and address missing values.
Step 3: Exploring data	Create data visualizations and calculate summary statistics to explore potential relationships in the dataset.
Step 4: Modeling data	Use modeling skills and content knowledge to fit and evaluate models, measure relationships, and make predictions.
Step 5: Interpreting data	Describe and interpret conclusions from data through written reports and presentations.

BYU

DEFINE THE PROBLEM

- What is the core problem?
- What processes, systems, orgs are affected?
- If solved, what is business value?
- How can problem be scoped?
- How is value measured?
- Characterize problem domain
- Is this a data-driven problem?
- What data is needed? (prelim)

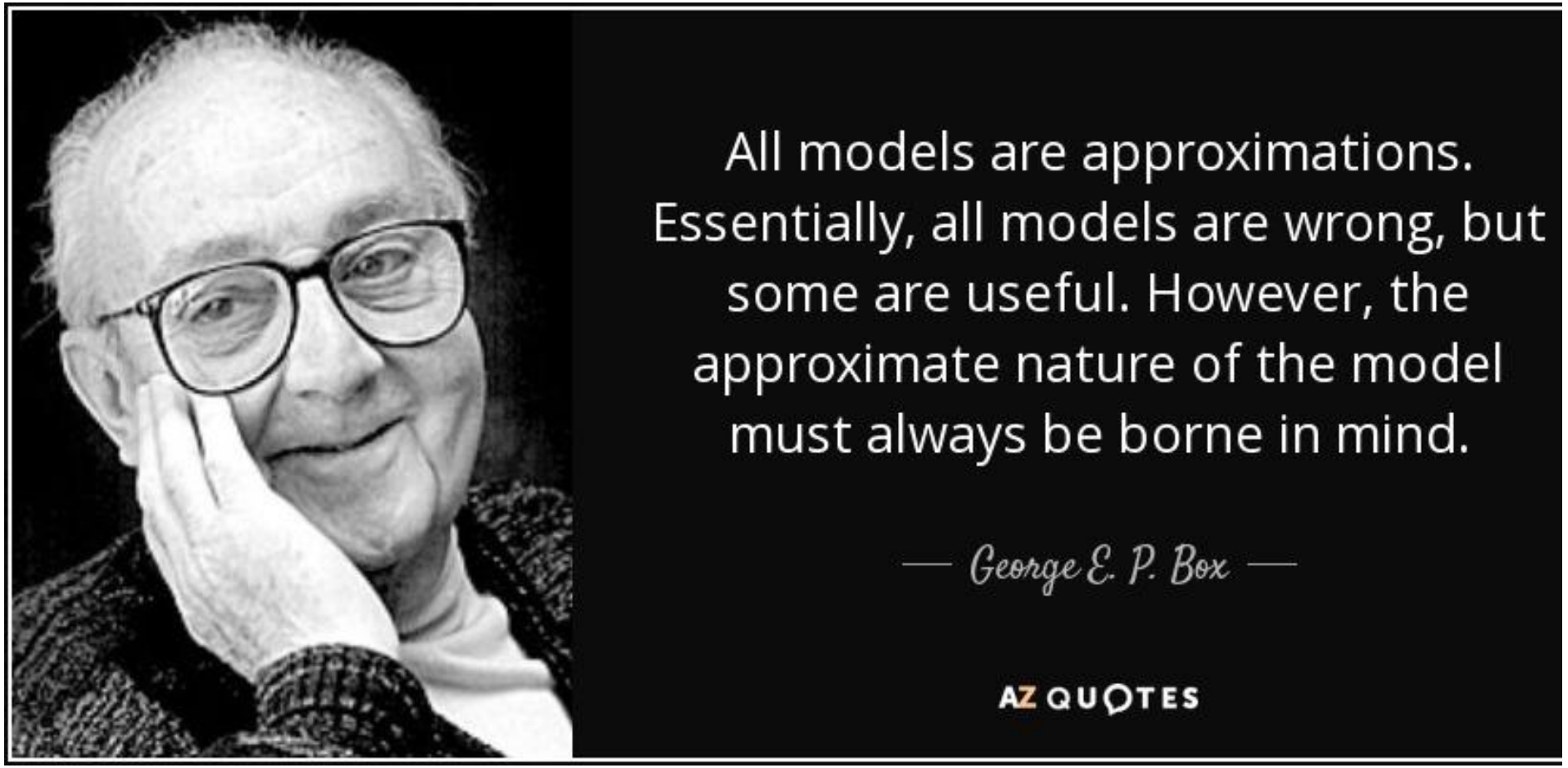
> IF I HAD AN HOUR TO SOLVE A PROBLEM I'D SPEND 55 MINUTES THINKING ABOUT THE PROBLEM AND 5 MINUTES THINKING ABOUT SOLUTIONS.



Albert Einstein

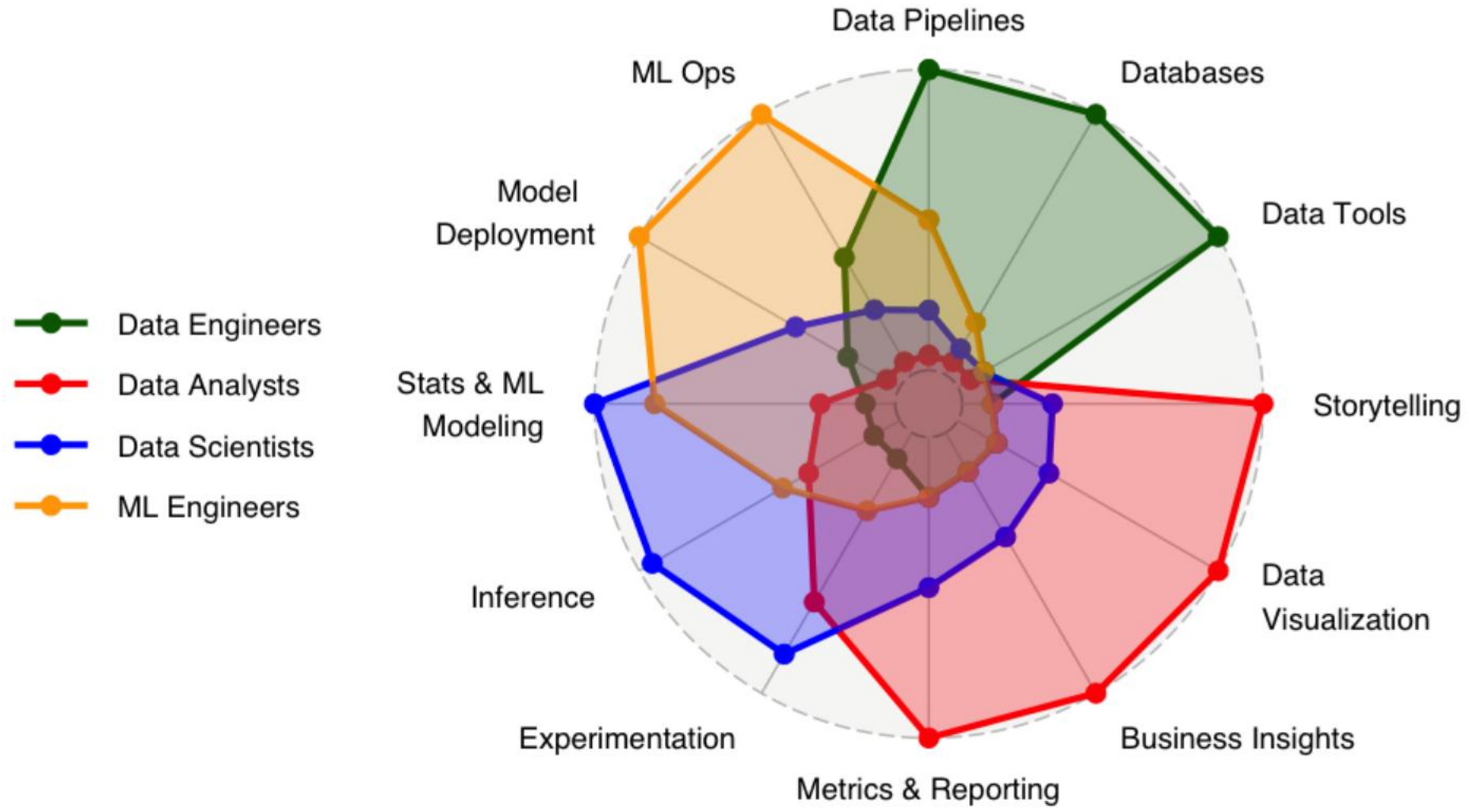
BYU

4. BUILD THE MODEL



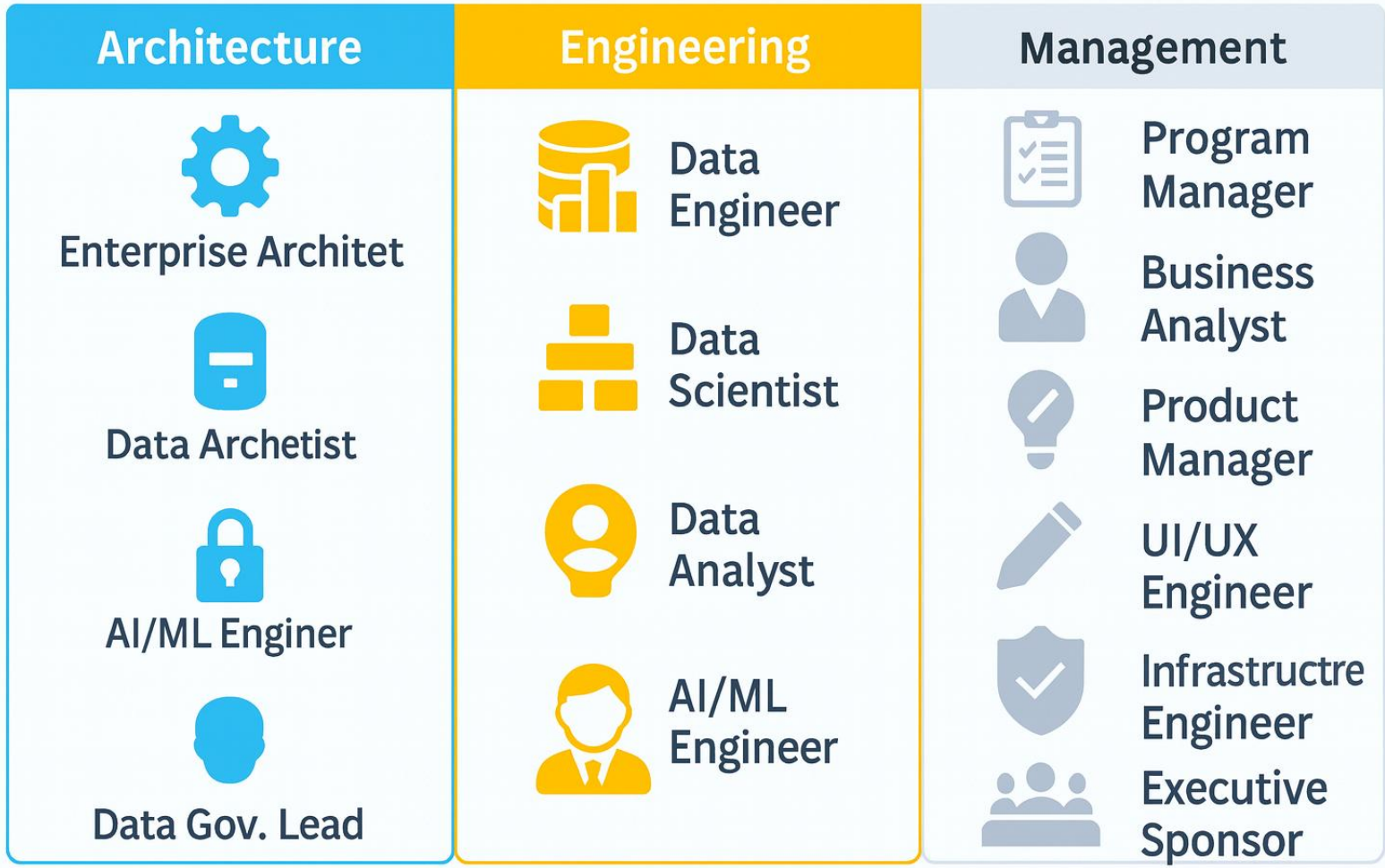
BYU

SPIDER CHART OF RELATIVE SKILLS FOR KEY DATA ROLES



From https://www.datacaptains.com/blog/guide-to-data-roles

ROLES ON A DATA SCIENCE APPLICATION PROJECT

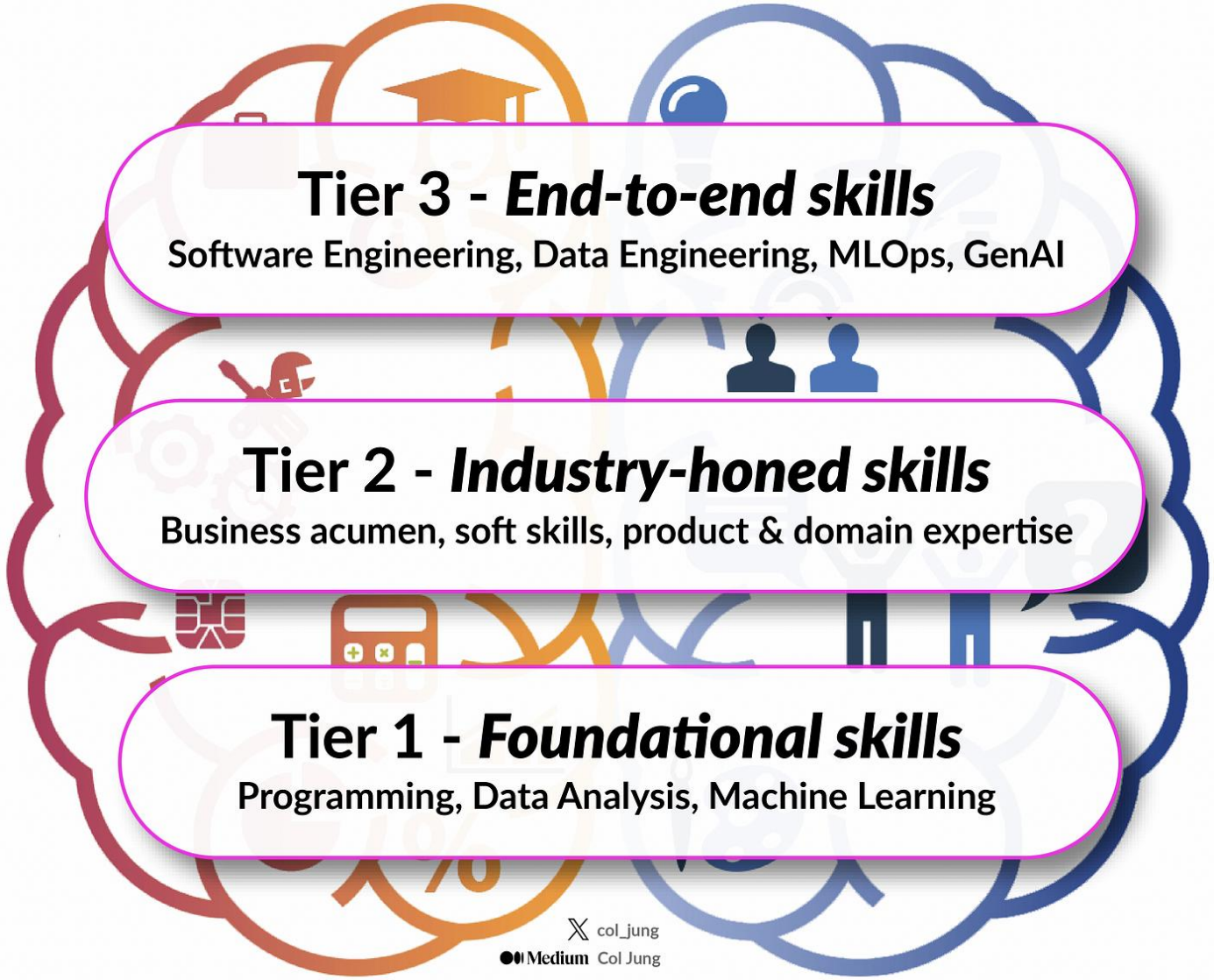


EXTENDED ROLES ON A DATA SCIENCE APPLICATION PROJECT

Role	Description
Data Engineer	Builds data pipelines, joins tables, converts data formats, prepares data for use by Data Scientists.
Data Scientist	Prepares data for modeling, extracts features, builds models
Data Analyst	Expert in SQL, BI, Excel, Analyzing data but not necessarily a domain expert (Tableau, PowerBI most popular tools)
AI/ML Engineer	Builds ML pipeline, integrate enterprise systems, monitor & manage models, skilled software engineer & data scientist
Enterprise Architect	Integrates DS applications into enterprise system (e.g., microservices, API gateway, event brokers, etc.)
Data Architect	Defines data management system architecture, data model
Data Governance Lead	Responsible for meta data, data catalog, data access, change management policies
Business Analyst	Expert in a particular domain (e.g., Finance), can use BI tools and Excel if set-up by the Data Analyst.
Program Manager	Keeps track of projects, personnel, budgets, identifies conflicts, dependencies, resource constraints
Application Engineer	Expert in technology for a particular domain or problem (e.g., Finance, Marketing, Sales, Manufacturing, etc.)
UI/UX Engineer	Designs, prototypes, and builds the user interface (mobile and web)
Product Manager	Responsible for overall product design, prioritization, deployment and assuring business value
Infrastructure Engineer	Expert in cloud and data management infrastructure
Security/Privacy Engineer	Assures application architecture is compliant with Enterprise security and privacy standards
Executive Sponsor	Oversees application development. Responsible for resourcing. Communicates with Executive Leadership.

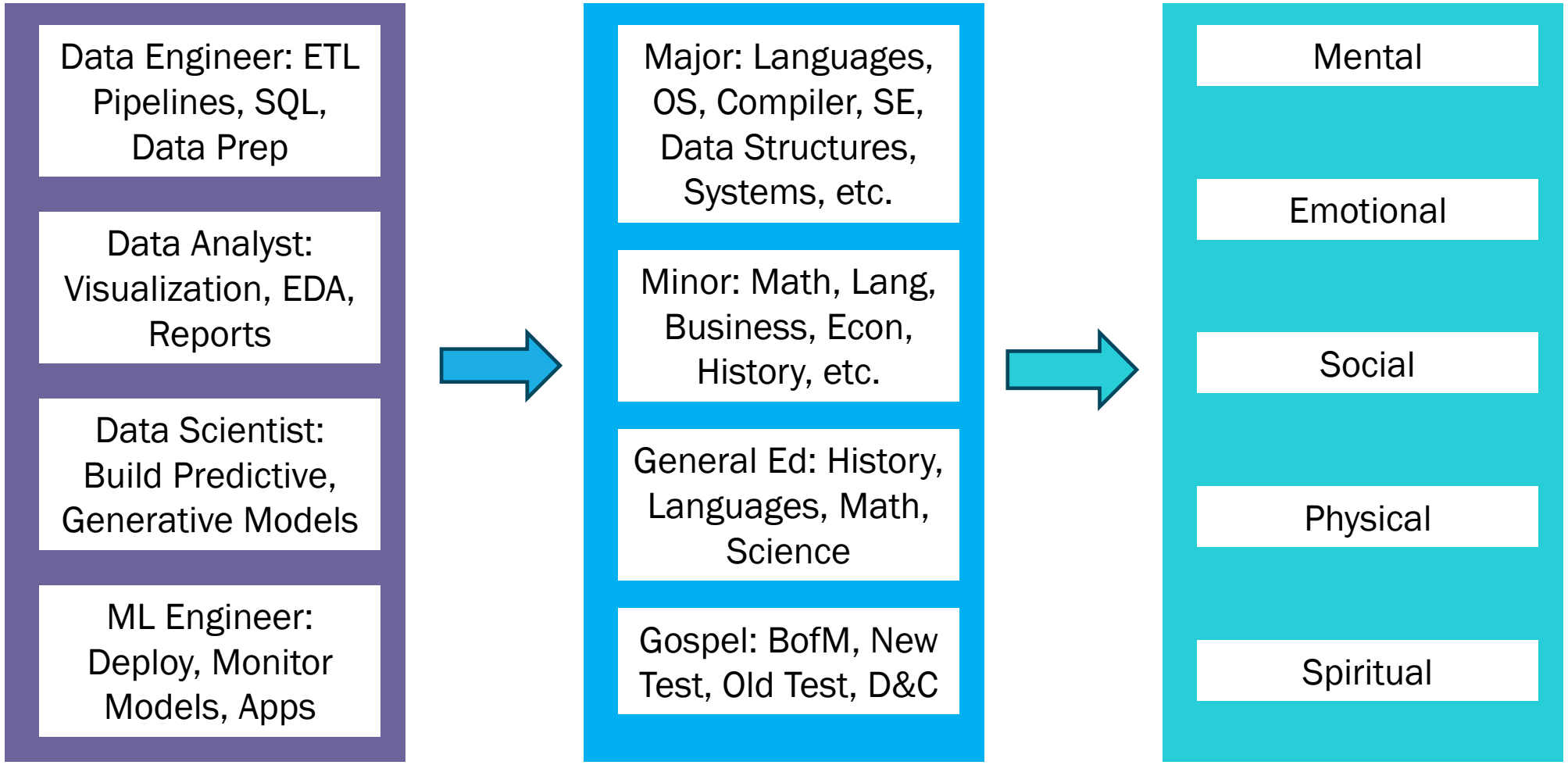
SKILL SETS FOR FULL STACK DATA SCIENTISTS

"Evolution of Data Science: New Age Skills for the Modern End-to-End Data Scientist," by Col Jung, Medium, July 23, 2024.

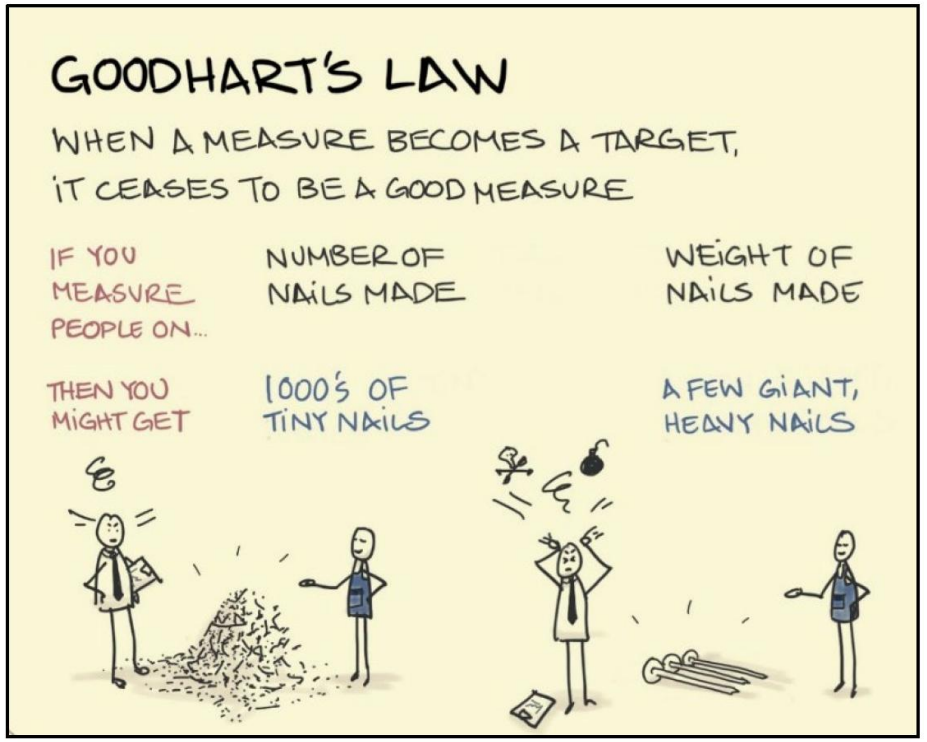


BYU

THE GOAL: FULL-STACK DATA SCIENTIST AND BEYOND



GOOD TO KNOW…



Students understand this law very well. It is easy to get caught up focusing on getting the "A" instead of mastering the knowledge or skill.

Don't let school… get in the way of your education ☺

BYU