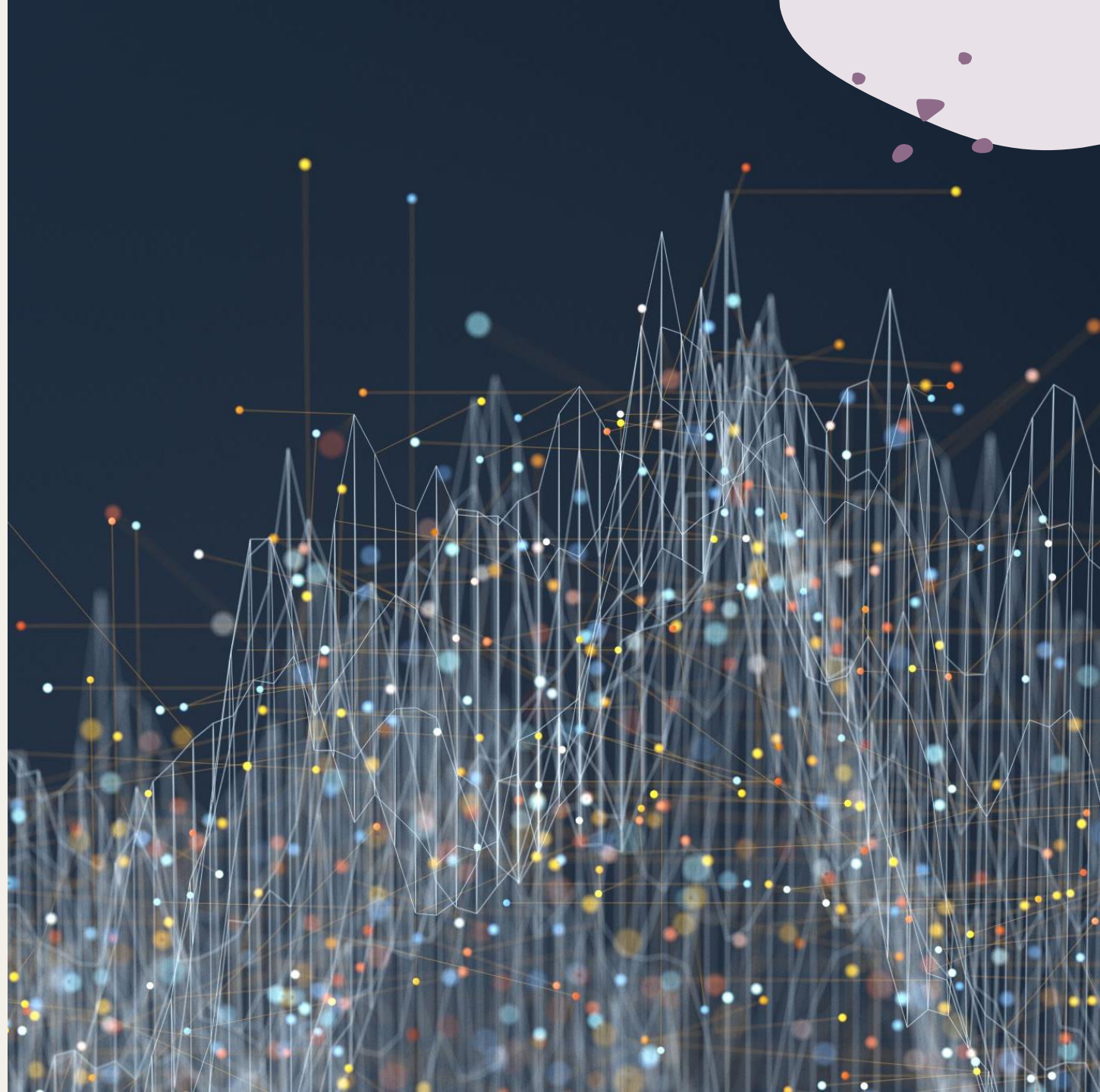


MMCi 411

Justin Waterfield, MMCi

Bioinformaticist

RTI International



Outline

Introduction

Before and After MMCi

Data Science Fundamentals

All of Us Research Program

Applications of Data Science

RTI International

Summary

Questions?



Introduction

Electron Microscopist at Duke University Health Systems for 11 years (2009-2020)

Graduated with a B.S. in Biology from Winthrop University

Part of the 2020 MMCi class, aka the "Covid Class"

TA for MMCi 517: Applied Data Science

Super nerd: board games, video games, and coffee fanatic

Love exploring Durham food scene with wife and doggo



Before and After MMCi!



Data Science Fun-da-mentals



Data Collection + Storage
[Data Engineer]



Data Preparation
[Data Scientist]



Exploration & Visualization
[Data Analyst]



Experimentation & Prediction
[Machine Learning Scientist]

All of Us Research Program

- [National Institutes of Health \(NIH\) – All of Us](#)
- NIH funded longitudinal study with 1M+ participants across the USA
- Focus on diversity and underserved populations
 - Rebuild medical trust
- Staging ground for precision medicine
 - Nutrition
 - Reducing costs (right treatment @ right time)
 - Environment + lifestyle + genetics
- Will create a huge open-access (to researchers) database that will help pave the way for future scientific endeavors

Data Science Examples and Applications

- Fintech
 - AI Advisor
 - Portfolio Manager
- Management/HR
 - Career Tracking
 - Hiring Keywords
- Supply Chain
 - Bots, bots, those damn bots!
 - Inventory Control
 - GPS + Shipping Routes
- Advertising/Marketing
 - Social Media
 - General Adversarial Networks (GAN) + Deepfakes
 - Algorithms + Consumer Preferences
- Healthcare
 - Digital Pathology
 - EHR + Machine Learning (NLP)
 - AI → Smart Diagnostics



RTI International

-
- Independent nonprofit research institute that utilizes science across many domains to improve the human condition
 - Founded in 1958, with HQ in Research Triangle Park (RTP)
 - Global workforce that has a progressive culture and focus on diversity-Majority of work comes from government grants/contracting
 - Bioinformatics area includes the [PhenX Toolkit](#): - A widely used phenotype + protocol search tool

Role of RTI

- Variable Mapping
 - Human Curation
- Machine Learning Training
- Data Processing + Cleaning
- Data Harmonization
 - Data Standards
- Identification of CDEs
(Common Data Elements)

- Example representation of surveys and EHR data in the *All of Us* database in the release of the registered tier in September 2019

Data Column in Database	EHR	Survey
observation_concept_id	4326027	43528559
(observation concept_name)	Family history of coronary arteriosclerosis	Diagnosed Health Condition: Father Circulatory Condition
value_as_concept_id	4167217	43528912
(value concept name)	Family history of clinical finding	Father Circulatory Condition: Heart Attack
observation_source_value	430091005 Family history of coronary arteriosclerosis (situation)	DiagnosedHealthCondition_FatherCirculatoryCondition
value_source_value	---	FatherCirculatoryCondition_HeartAttack

Wrap-up

- Hang in there, it is a marathon
- Remember the fun-da-mentals!
 - Import/Collect → Store/Prepare → Visualize → Experiment
 - Watch out for biases
 - Don't get stuck, experiment!
- All of Us = Kind of a big deal
- Data Science is all around us...
- Be thinking of your future and career aspirations
- Start learning to code!!!!
 - Python + R



Questions?



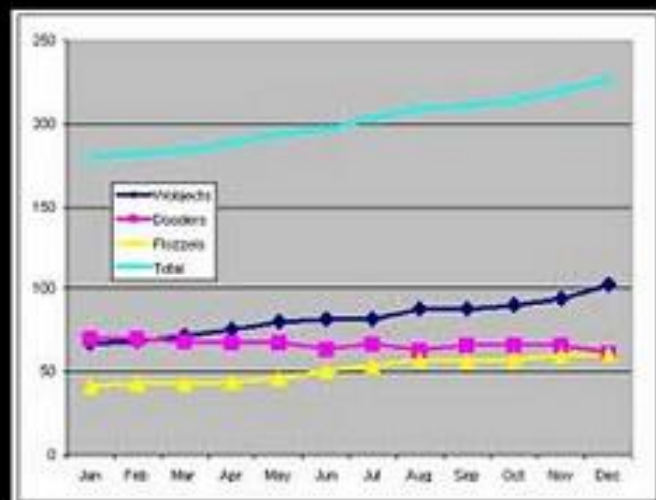
What my friends think I do



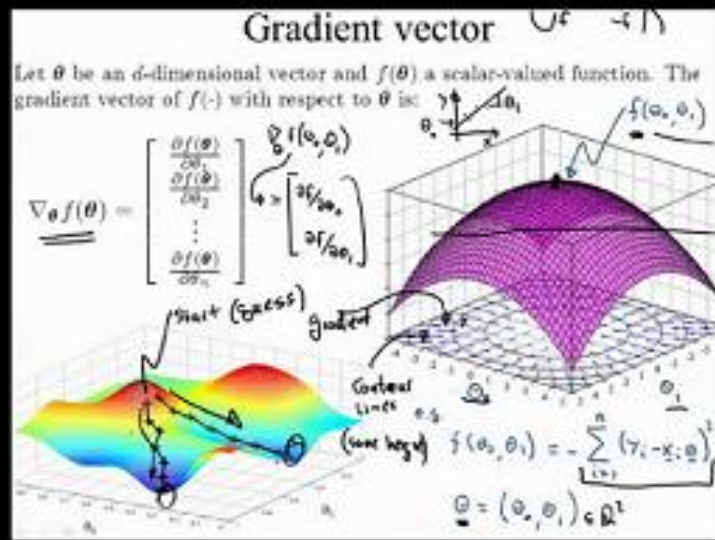
What my mom thinks I do



What society thinks I do



What my boss thinks I do



What I think I do



What I actually do