

TechMoms: Data Science

Jessica Langford | Adobe | 09.24.2022

Agenda



01 INTRODUCTION

My professional, educational, and personal background

02 WHAT IS DATA SCIENCE AND WHY DO WE NEED IT?

Increased complexities of data and technologies created need for Data Science

03 MY JOB AS A DATA SCIENTIST

What do I actually do everyday??

04 PATHWAYS TO BECOMING A DATA SCIENTIST

How you can work towards becoming a Data Scientist

05 REAL WORLD EXAMPLE IN R

Code!



Introduction: About Me

PROFESSIONAL

Data Scientist, Adobe

Adobe Digital Insights, Adobe
August 2017 - Present

Data Science Consultant

Adobe Consulting Services, Adobe
May 2011 – July 2017

Actuarial Analyst

Defined Benefits, Watson Wyatt
September 2007 – October 2008

EDUCATIONAL

Bachelors of Science

Actuarial Science

Economics

Brigham Young University, Provo
August 2002 – April 2007

Master of Statistics

Brigham Young University, Provo
August 2009 – August 2012



 Email: jessica.lyn.Langford@gmail.com

 Phone: 410-507-4955

 Location: Eagle Mountain, UT

 LinkedIn: <https://www.linkedin.com/in/jesslangford/>

Introduction: About Me (cont.)

PERSONAL

- Grew up in Crofton, Maryland
- Played basketball for Dixie State University
- Big fan of the rock band The Who
- Met my husband, Taylor, the same day I started working at Adobe and we were married 6 months later ☺
- Jackson – 6.5
- Emi (short for Emerson) – 4
- Enjoy watching sports, watching trashy reality TV shows, hiking, snowboarding, hanging with family and friends
- Current fixation is building an STR in Tollgate Canyon



What is Data Science and Why do we Need it?

The New Oil

"Big data is the new oil. The folks who are going to get good value are those who are going to be able to refine it and turn it into useful products."

Andy Cutler, Director of Strategy, SAS

- Essential Resource
- Value must be extracted
- Complex processing



Why do we need Data Science?

GOOD DATA > OPINION

MORE DATA \neq BETTER
INFORMATION

DATA SIZE

DATA COMPLEXITY



Why do we need Data Science?

GOOD DATA > OPINION

MORE DATA \neq BETTER
INFORMATION

DATA SIZE

DATA COMPLEXITY

DOLLARS PER WIN



\$2.1M
DOLLARS PER WIN

The Oakland Athletics logo, featuring the stylized 'A's' in green and gold.
\$678K
DOLLARS PER WIN

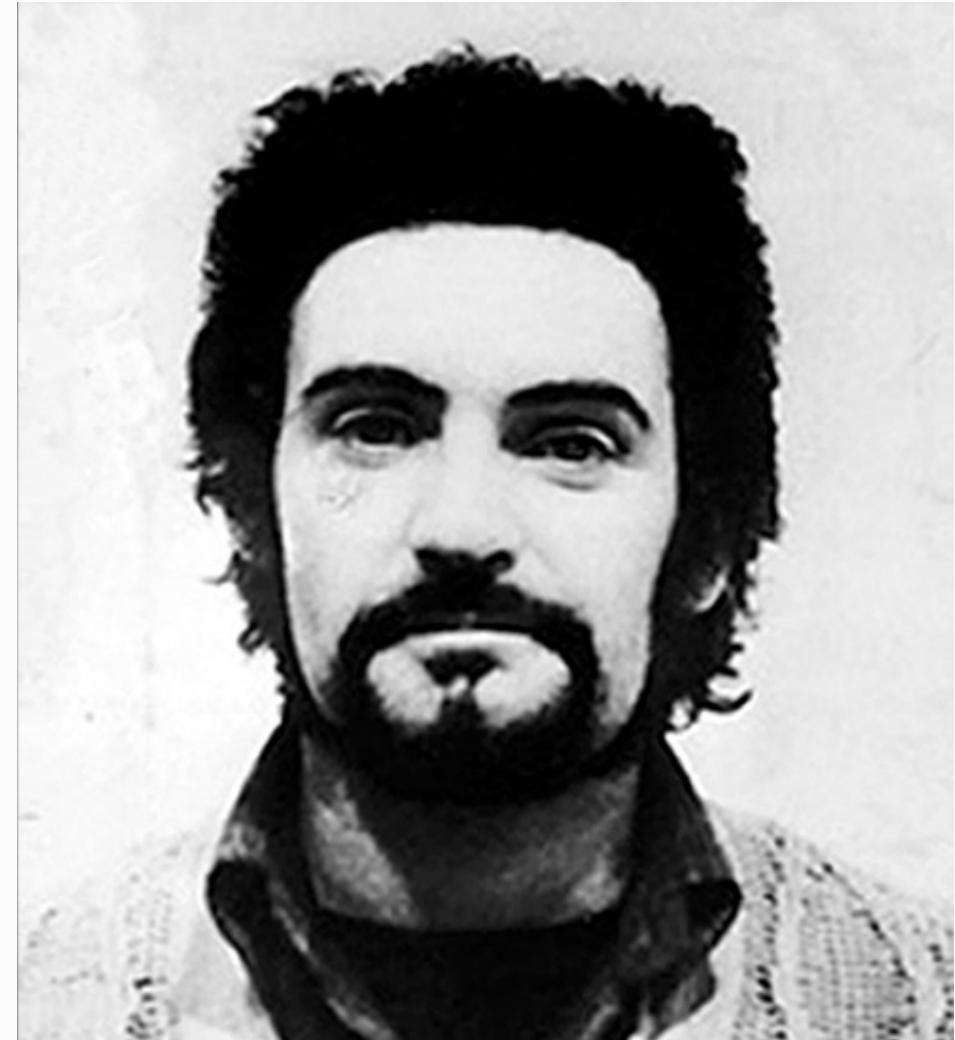
Why do we need Data Science?

GOOD DATA > OPINION

MORE DATA \neq BETTER
INFORMATION

DATA SIZE

DATA COMPLEXITY



Why do we need Data Science?

GOOD DATA > OPINION

MORE DATA ≠ BETTER

INFORMATION

DATA SIZE

DATA COMPLEXITY



Why do we need Data Science?

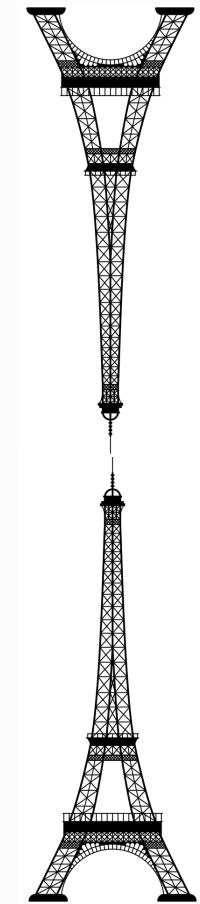
GOOD DATA > OPINION

MORE DATA \neq BETTER
INFORMATION

DATA SIZE

DATA COMPLEXITY

2,500,000,000,000,000



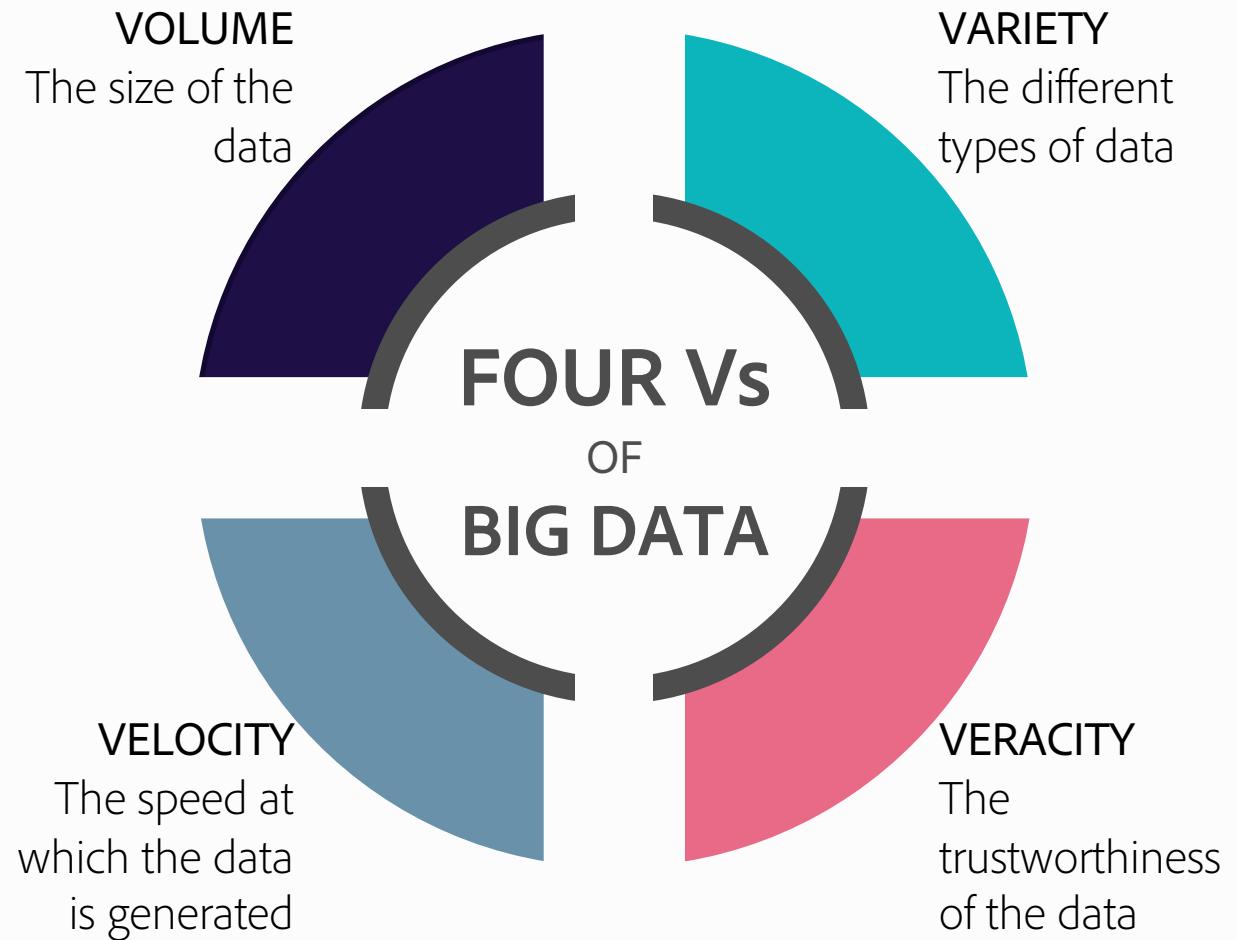
Why do we need Data Science?

GOOD DATA > OPINION

MORE DATA ≠ BETTER INFORMATION

DATA SIZE

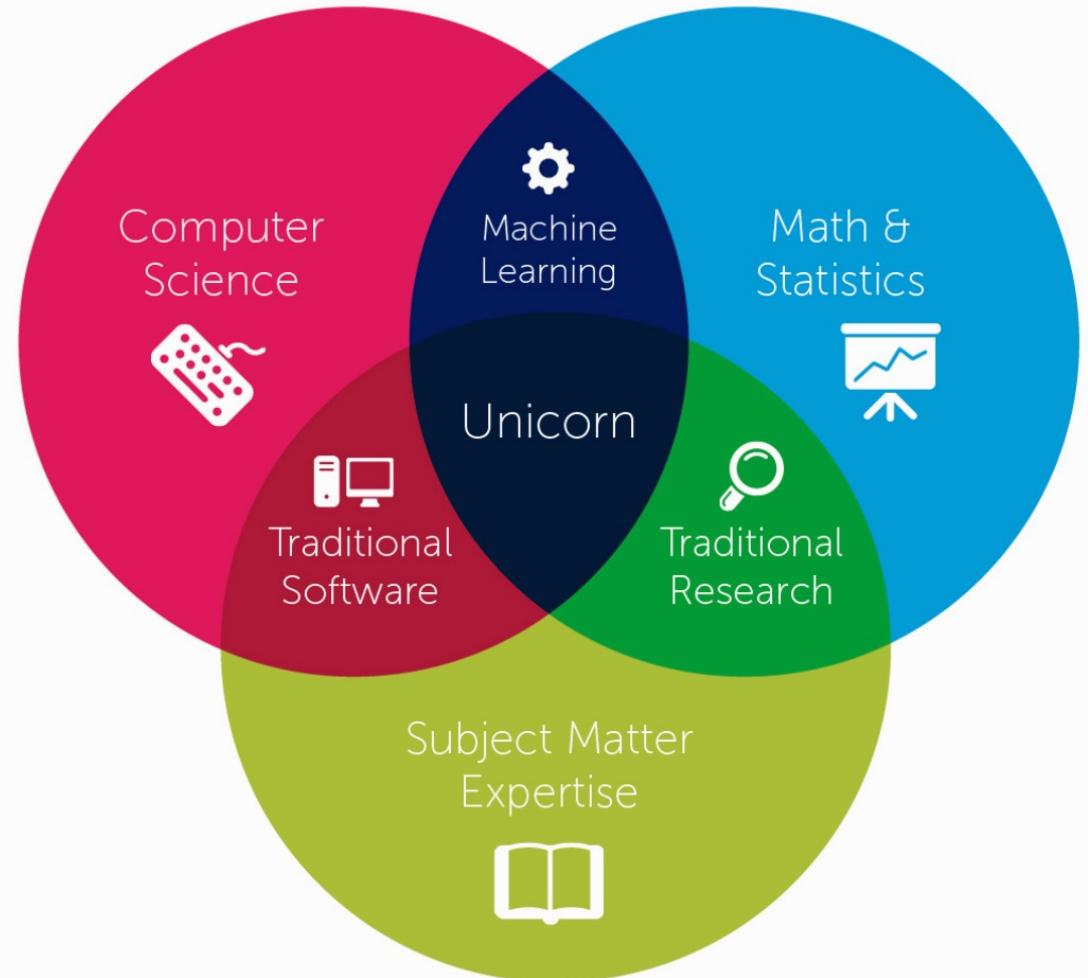
DATA COMPLEXITY



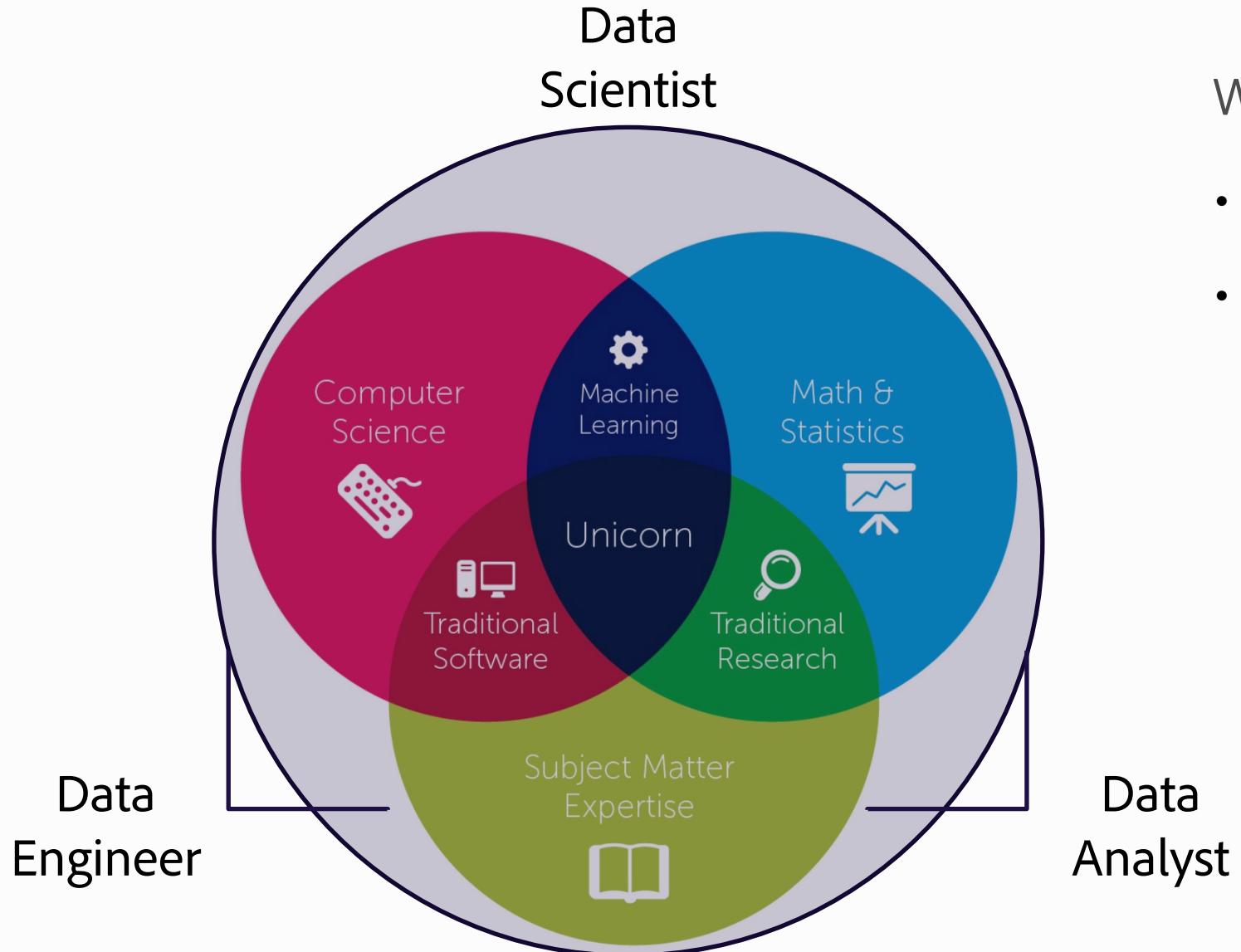
What is Data Science?

Data Science is a multi-disciplinary field that uses scientific methods, processes, algorithms, and systems to **extract knowledge and insights** from structured and unstructured data.

Data scientists are a new breed of **analytical data expert** who have the **technical skills** to solve complex problems – and the **curiosity** to explore what problems need to be solved.



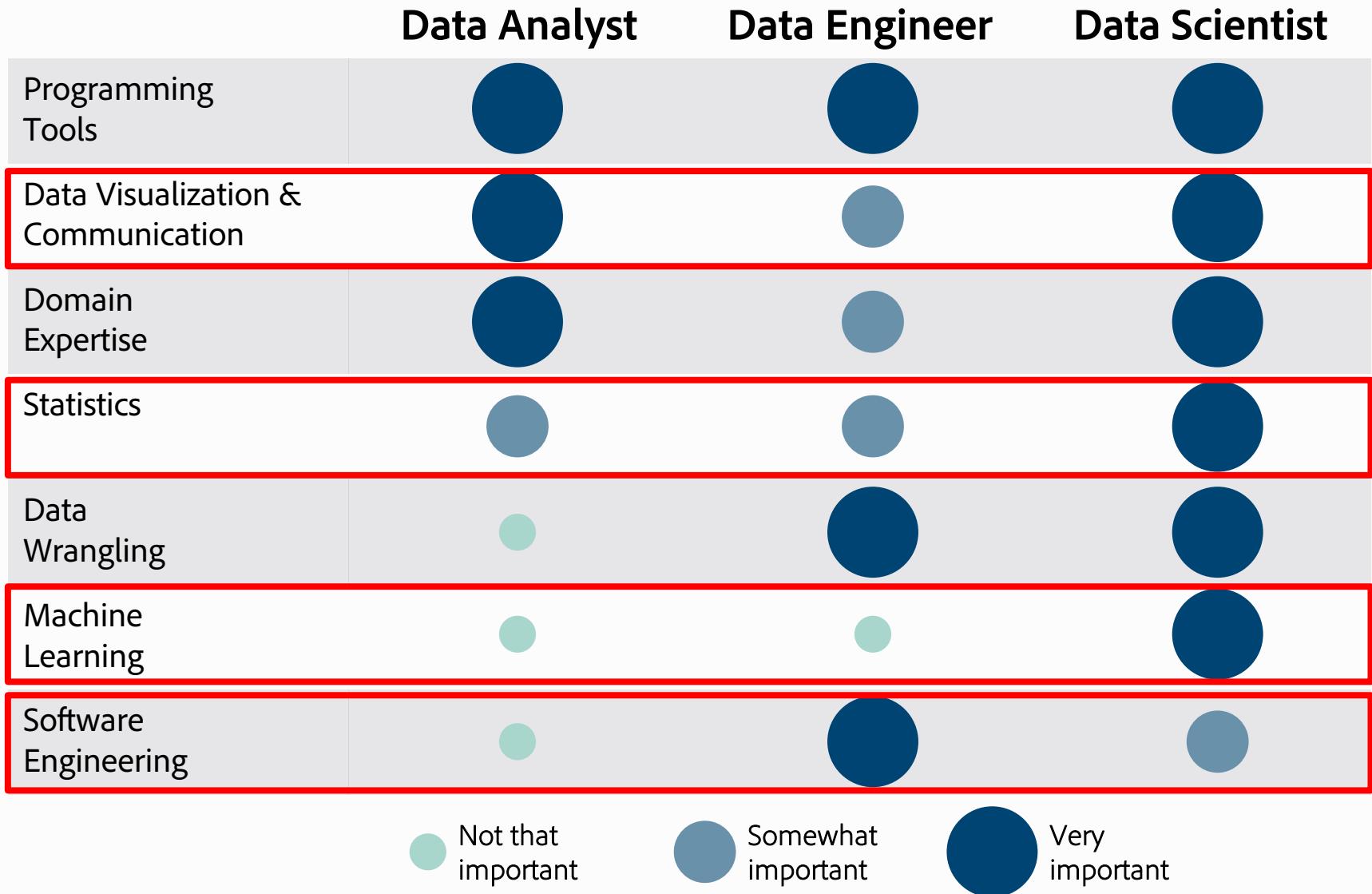
Comparing Data-Related Careers



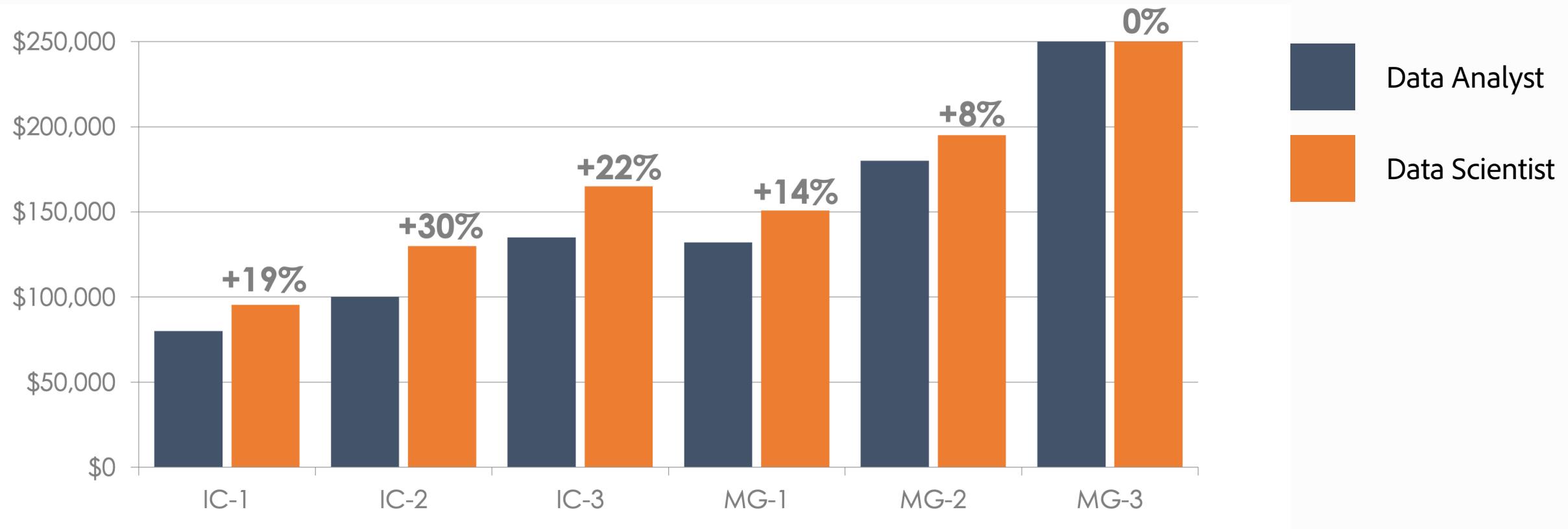
WHERE DO YOUR INTERESTS LIE?

- Organizations have needs for all three of these data-related job roles
- Many Data Scientists start as either a Data Engineer or Data Analyst and learn a third discipline over time

Comparing Data-Related Careers: Necessary Skills



Comparing Data-Related Careers: Salaries



data © 2020 Burtch Works LLC

IC = Individual Contributor
Level 1 = 1-3 YOE
Level 2 = 4-8 YOE
Level 3 = 9+ YOE

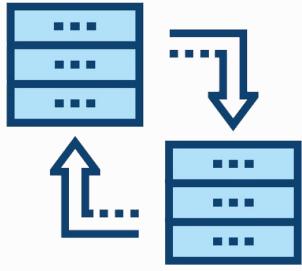
MG = Manager
Level 1 = 1-3 DR
Level 2 = 4-9 DR
Level 3 = 10+ DR

Data Science Process



Business Problem

Deep understanding of the problem you're trying to solve or the question you're trying to answer



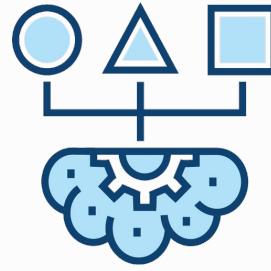
Understanding Data

Information regarding how the data was collected



Data Preparation

Understanding how the data is stored, how to access it, joining with other data sets, etc.



Modeling & Analysis

Applying combinations of statistical algorithms and machine learning methods



Sharing Results

Communicating the results to stakeholders clearly and efficiently



Deployment

Creating a process to do everything in a fast, repeatable, and automated way

Data Scientist Tools

Store Data & ETL



Exploratory Analysis



Model Building & Insights



Visualization



Model Execution in Production



WHAT STACK IS BEST?

- May depend on the type of data and industry you're in
- May vary from project to project
- Most important skill is to learn to learn
 - Stackoverflow
 - Product documentation
 - Google searches



My Job as a Data Scientist

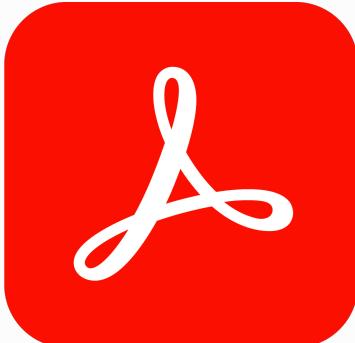
Adobe Product Overview



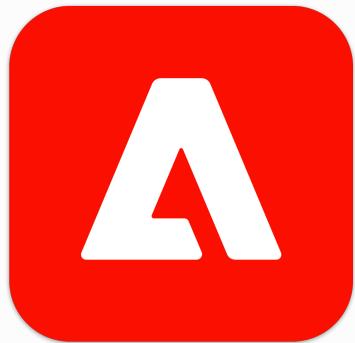
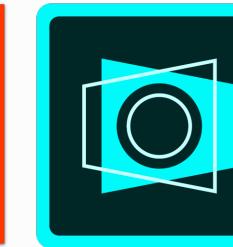
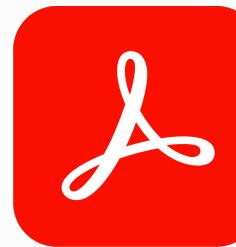
ADOBE CREATIVE CLOUD



+



ADOBE DOCUMENT CLOUD

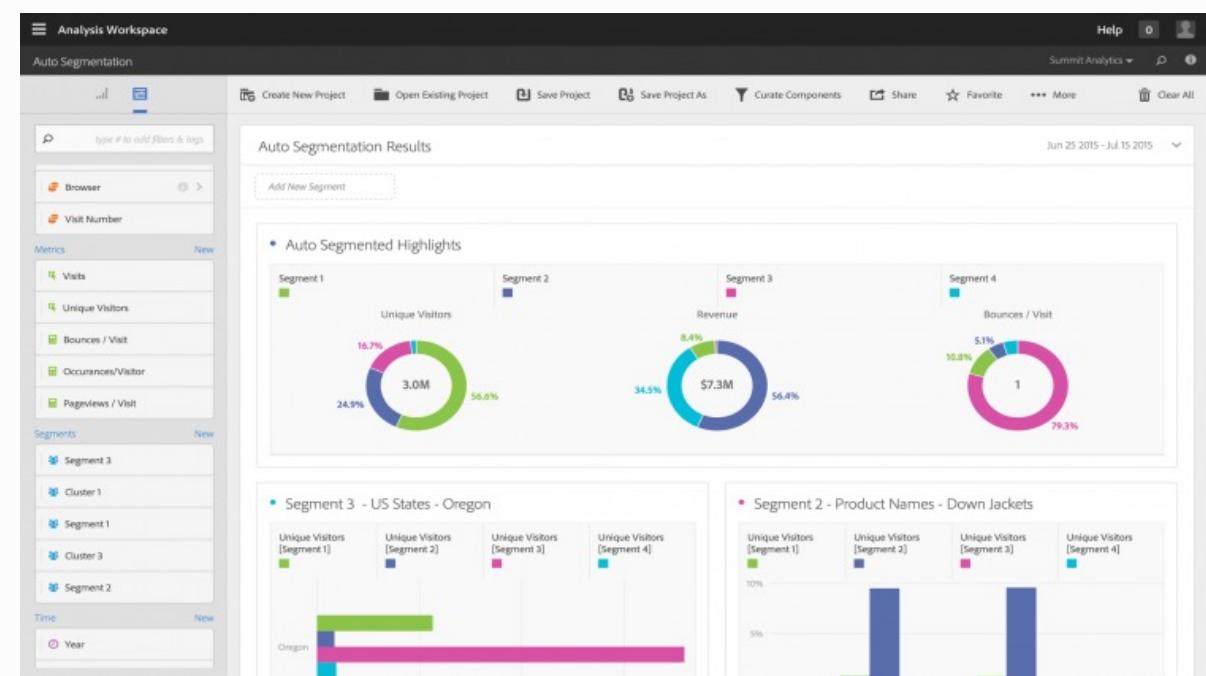
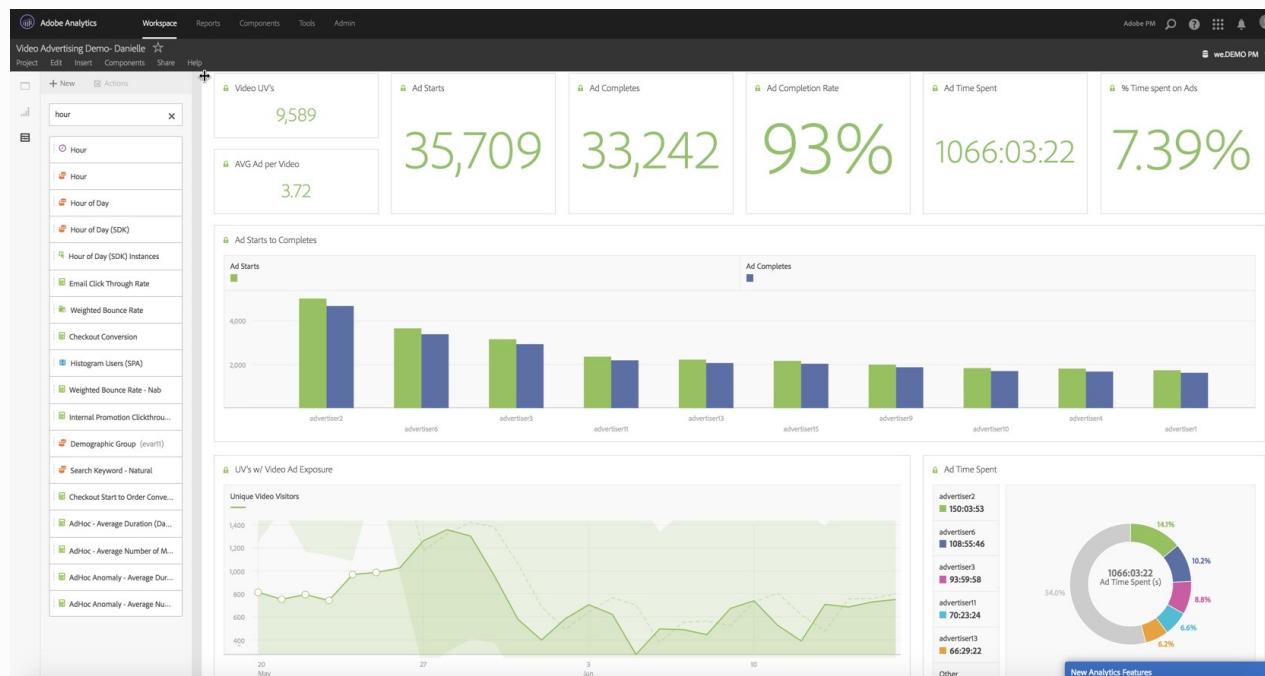
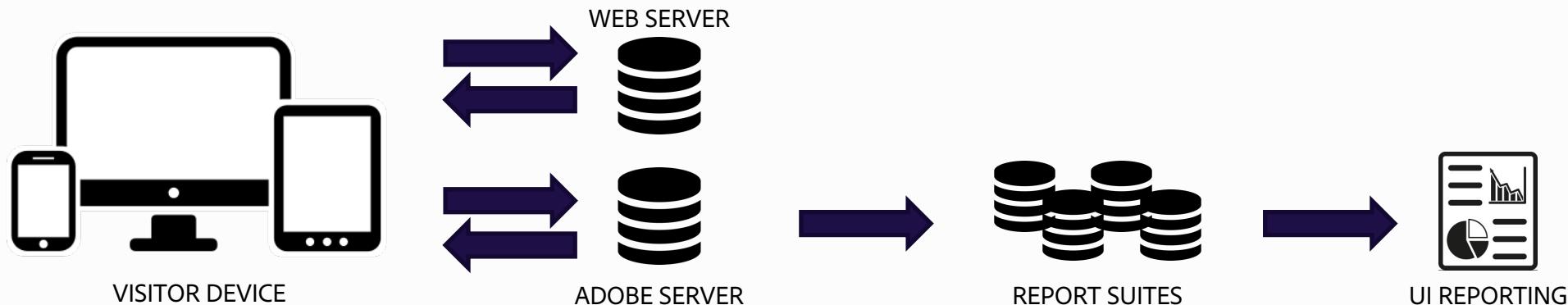


ADOBE EXPERIENCE CLOUD



+

Adobe Analytics Data Overview



Adobe Analytics Data Overview

TRILLIONS

BILLIONS

TRANSACTIONS

MILLIONS

UNIQUE VISITORS

HITS

Adobe Digital Insights: DEI

DIGITAL ECONOMY INDEX



- Dynamic measure of commerce among the products and services consumers are buying online
- Partner with leading experts in government and academia
- Provides a more complete and real-time picture of the global economy and inflation
- Spans industries including Travel, Retail, Financial Services, etc.
- Very useful to businesses and public organizations in understanding global economic conditions

Adobe Digital Insights: Holiday Report

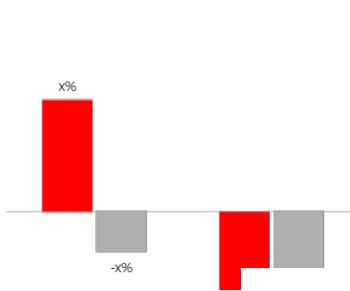
HOLIDAY REPORT

- For many retailers, 9 of the top 10 shopping days of the year fall between Thanksgiving and Christmas
- Based on historical data, we can accurately model eCommerce trends for the holiday season
- Extremely important to helping businesses plan, prepare, and stay competitive

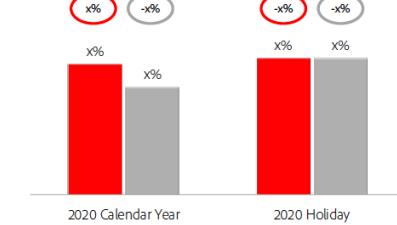
Adobe Digital Insights: Benchmark Reports

A1 Overall Performance - Annual & Holiday YoY Growth

Digital Revenue YoY % Change¹



Digital Conversion Rate³



[Client] [Peer Set]
YoY % Change
0

Key Takeaways

- Point 1
 - Sub-point
- Point 1
 - Sub-point

A1.4 Thanksgiving Week Performance by Hour

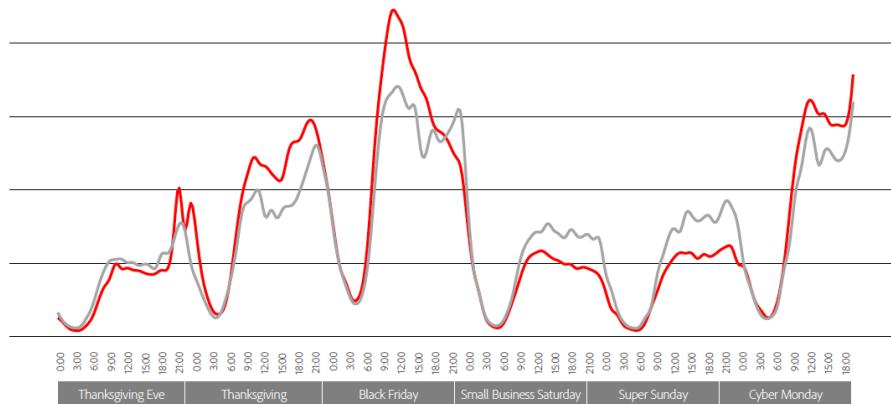
Hourly Revenue Share (%) | Thanksgiving Week as 100%¹

2020 Calendar Year

2

1. Calendar Year includes date range from Jan 1st to December 31st.
2. Holiday is defined as October 1st to December 31st.
3. Conversion Rate = Orders/Visits

© 2020 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.



1. Thanksgiving week accounts for 25% of total holiday revenue.
2. Hours are adjusted to EST time zone.

© 2020 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.



BENCHMARK REPORTS

- Adobe's clients are often interested in how they are performing relative to their peers
- We have the ability to do analysis on the most important Key Performance Indicators (KPIs)
- Help businesses stay competitive and understand their digital capabilities
- Viewed as a strategic tool for building relationships with clients
- Reports are used by CSMs, AEs, and pre-sales consultants to help in the sales and value proposition to clients

Adobe Digital Insights: Challenges



BIG DATA

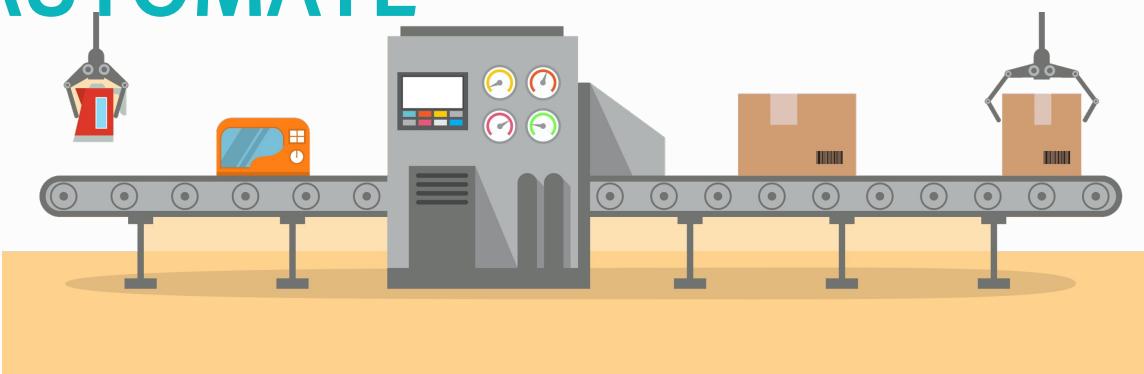


UNORGANIZED DATA

COMPLEX MODELING

A dense collection of mathematical equations, graphs, and diagrams. It includes a complex integral equation, a 2D plot with shaded regions, a matrix multiplication, and various mathematical symbols like α , β , γ , δ , and π . The equations involve variables x and y , and constants like π and e .

AUTOMATE DATA



Things I Like About My Job

- Work on Interesting Problems
- Making an Impact on Adobe's Bottom Line
- Get to Be Creative
- Opportunities to Learn and Increase Skill Set
- Adobe/Tech Culture
- Fun Teammates
- My Sanity

Things I Like About My Job

- Work on Interesting Problems
- Making an Impact on Adobe's Bottom Line
- Get to Be Creative
- Opportunities to Learn and Increase Skill Set
- Adobe/Tech Culture
- Fun Teammates
- My Sanity



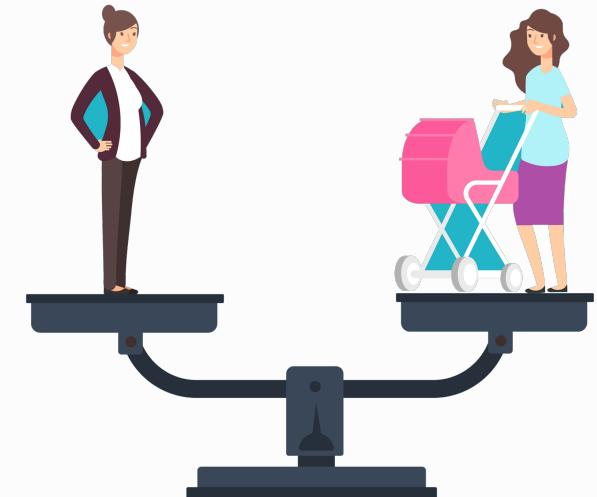
Things I Like About My Job

- Work on Interesting Problems
- Making an Impact on Adobe's Bottom Line
- Get to Be Creative
- Opportunities to Learn and Increase Skill Set
- Adobe/Tech Culture
- Fun Teammates
- My Sanity



Things I Like About My Job

- Work on Interesting Problems
- Making an Impact on Adobe's Bottom Line
- Get to Be Creative
- Opportunities to Learn and Increase Skill Set
- Adobe/Tech Culture
- Fun Teammates
- My Sanity



Why I Have No Intention of Leaving My Job



- Taylor took a “little” extra time to finish his education
- Started his undergraduate degree shortly after we were married in 2012
- Graduated with his PharmD and MBA degrees in May 2020
- **I was able to support my husband and provide for my family on a single income**

- Jackson has had speech and other social developmental delays since early childhood
- Official ASD diagnosis June 2020
- Since a very young age, he has been many programs to help him overcome some of his unique challenges
- **Medical benefits include ABA therapy, speech therapy, occupational therapy, etc.**

- My husband and I struggled for several years of infertility
- Emi and Jackson are both the product of IVF
- **Medical coverage made it financially reasonable to pursue advanced reproductive technologies**

Pathways to Becoming a Data Scientist

Pathways to Data Science: Formal Education

FULL TIME PROGRAMS (UTAH BASED)

- BYU
 - Undergraduate: [BS Data Science](#), [BS Computer Science](#)
 - Graduate: [MS Statistics](#), [MS Computer Science](#)
- Southern Utah University:
 - Graduate: [MS Business Analytics](#)
- University of Utah:
 - Undergraduate: [BS Data Science](#), [BS Computer Science](#)
 - Graduate: [MS Business Analytics](#)
- Utah State University:
 - Undergraduate: [BS Statistics](#), [BS Computer Science](#)
 - Graduate: [MS Data Science](#), [MS Computer Science](#)
- Utah Valley University:
 - Undergraduate: [BS Statistics](#)
- Westminster:
 - Undergraduate: [BS Computer Science](#), [Data Science Minor](#)

- Weber State University:

- Undergraduate: [BS Computational Statistics and Data Science](#), [BS Computer Science](#)
- Graduate: [MS Computer Science](#)

ONLNE & PART TIME PROGRAMS

- ["Best" Online BS Data Science Programs](#)
- ["Best" online MS Data Science Programs](#)

THINGS TO CONSIDER

- What type of learner are you?
- Online vs In-class experience
- Specific concentrations or areas of interest
- Cost
- Time commitment
- Pre-requisites (courses, tests, etc.)

Pathways to Data Science: Informal Education

COURSES & CERTIFICATES

- Udemy:
 - [Complete Data Science Bootcamp \(Python\)](#)
 - [Data Science and Machine Learning Bootcamp with R](#)
- Coursera:
 - [Johns Hopkins Data Science Specialization](#)
- DataCamp:
 - [Data Science with R](#)
 - [Data Science with Python](#)
- Udacity:
 - [Data Science Nanodegree](#)
- edX:
 - [Certificates, MicroBachelors, MicroMasters](#)

BOOTCAMPS

- [Galvanize](#)
- [General Assembly](#)

THINGS TO CONSIDER

- Pricing Structure
 - Monthly Fee vs Flat Fee
- Time commitment
 - Go at Your Own Pace vs Regular Class Times
- Pre-requisites

Pathways to Data Science: DIY

FREE/INEXPENSIVE TRAINING RESOURCES

- Find a Shadow/Mentor
- People to Follow on Social Media (mostly on Twitter)
 - Hadley Wickham
 - Nate Silver
- Data Science Blogs
 - [FiveThirtyEight](#)
 - [Data Science Central](#)
 - [Kdnuggets](#)
 - [RStudio](#)
- LinkedIn Professional Groups
- Books
 - [An Introduction to Statistical Learning](#)
 - [R for Data Science](#)
 - [Python Data Science Handbook](#)
 - [Doing Data Science](#)

EXPERIENCE

- Kaggle Competitions/Projects
 - <https://www.kaggle.com/>
- Start your own GITHUB repository
 - <https://docs.github.com/en/github/getting-started-with-github>
- Find and analyze a data set that interests you
 - Google is your friend!
- Create a Portfolio

Real World Example in R

Hands On Example: Download & Install R

WHAT IS R?

- R is a free software environment for statistical computing and graphics
- Full-featured statistical analysis package that allows you to:
 - Import and export data in many formats
 - Manipulate & analyse data
 - Apply statistical models to data
 - Graphically represent data
- Comes pre-loaded with a lot of functionality, but additional functionality can be installed (called packages or libraries)

WHAT IS RSTUDIO?

- Integrated development environment (IDE) for R
- Universally considered the best interface for R programming

INSTALL BASE R

- <http://www.r-project.org>

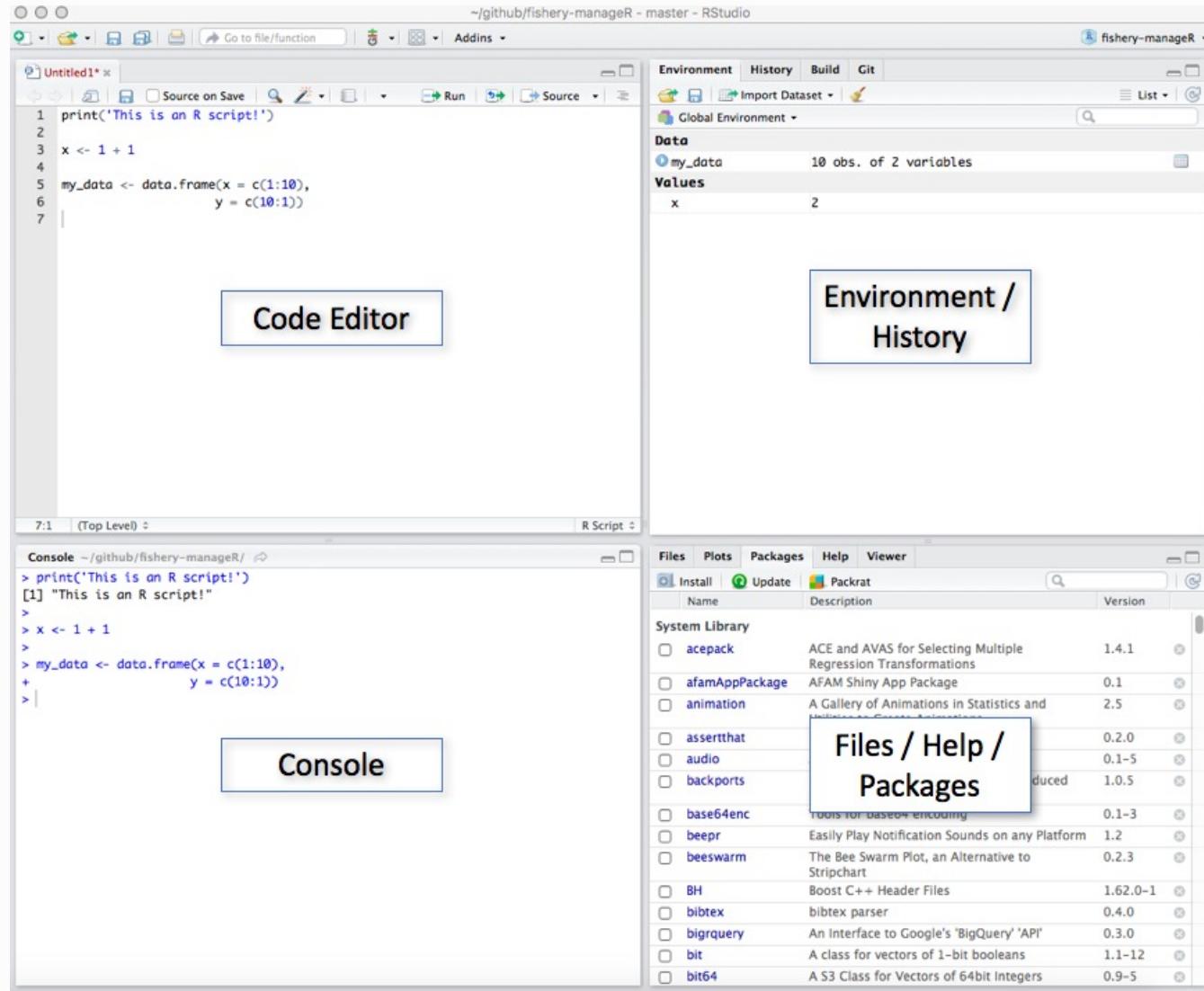
INSTALL RSTUDIO

- <http://www.rstudio.com>

INSTALL ADDITIONAL R PACKAGES

- `install.packages("packagename")`

Hands On Example: RStudio Interface



CODE EDITOR

- Where you write your code (often called a 'script') for your analysis
- Each tab represents a different R script (.R) file

CONSOLE

- Where R prints the output of the code you're running
- You can run code from your code editor or execute code directly in the console

ENVIRONMENT/HISTORY

- Environment – Displays all your data, variables, and user-defined functions
- History – A list of your code/command history

FILES/HELP/PACKAGES

- Files – List of all the files contained in your working directory
- Plots – Commands related to plots print here
- Packages – List of packages currently installed on your computer
- Help – R documentation on functions

Hands On Example: Getting Started in RStudio

GITHUB REPO:

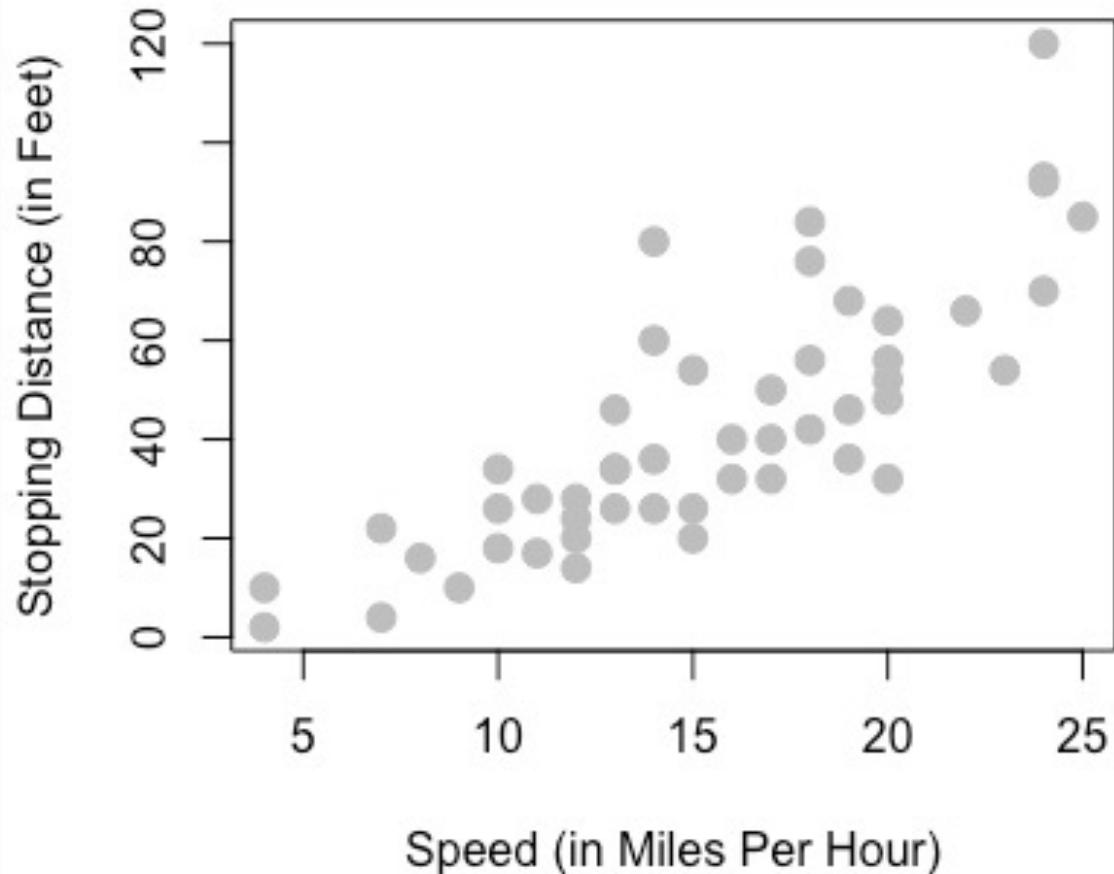
https://github.com/jlynlangford/tech_moms

FILE NAME:

my_first_R_script.R

Hands On Example: Simple Linear Regression

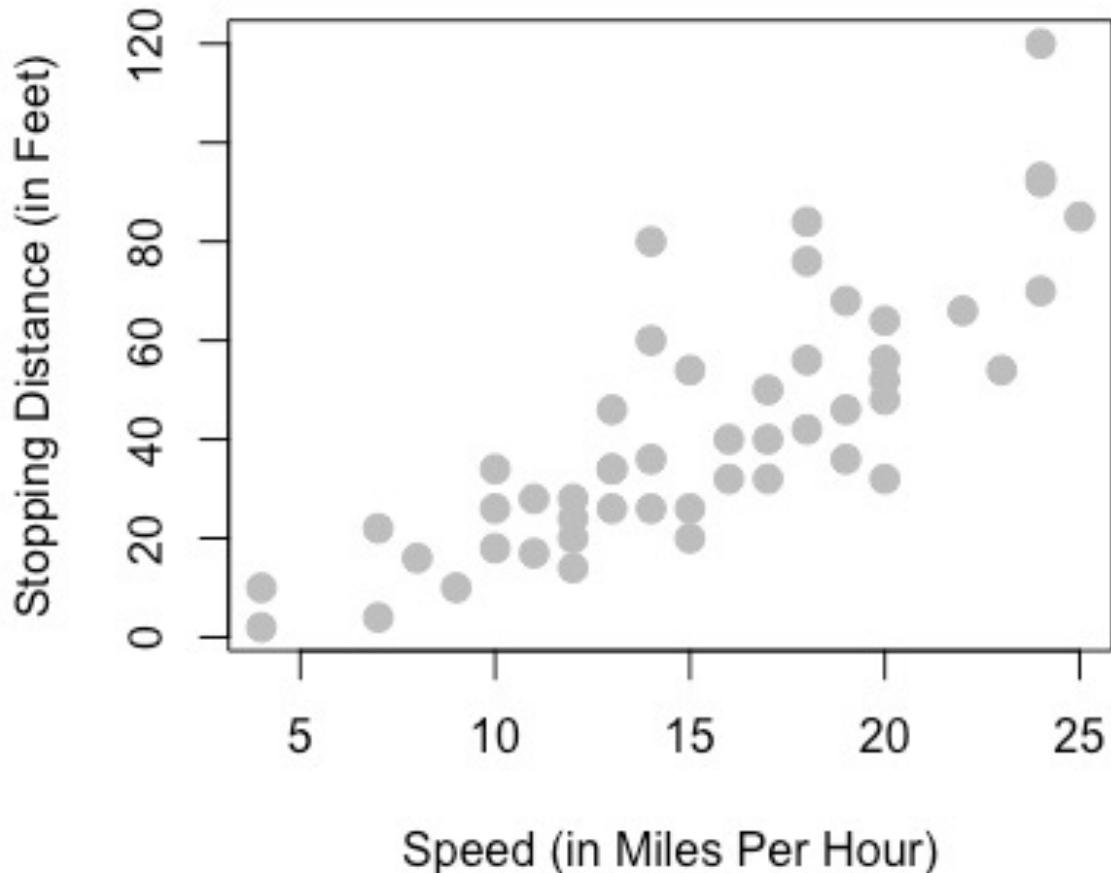
Stopping Distance vs Speed



- How would you describe the relationship between Stopping Distance and Speed?

Hands On Example: Simple Linear Regression

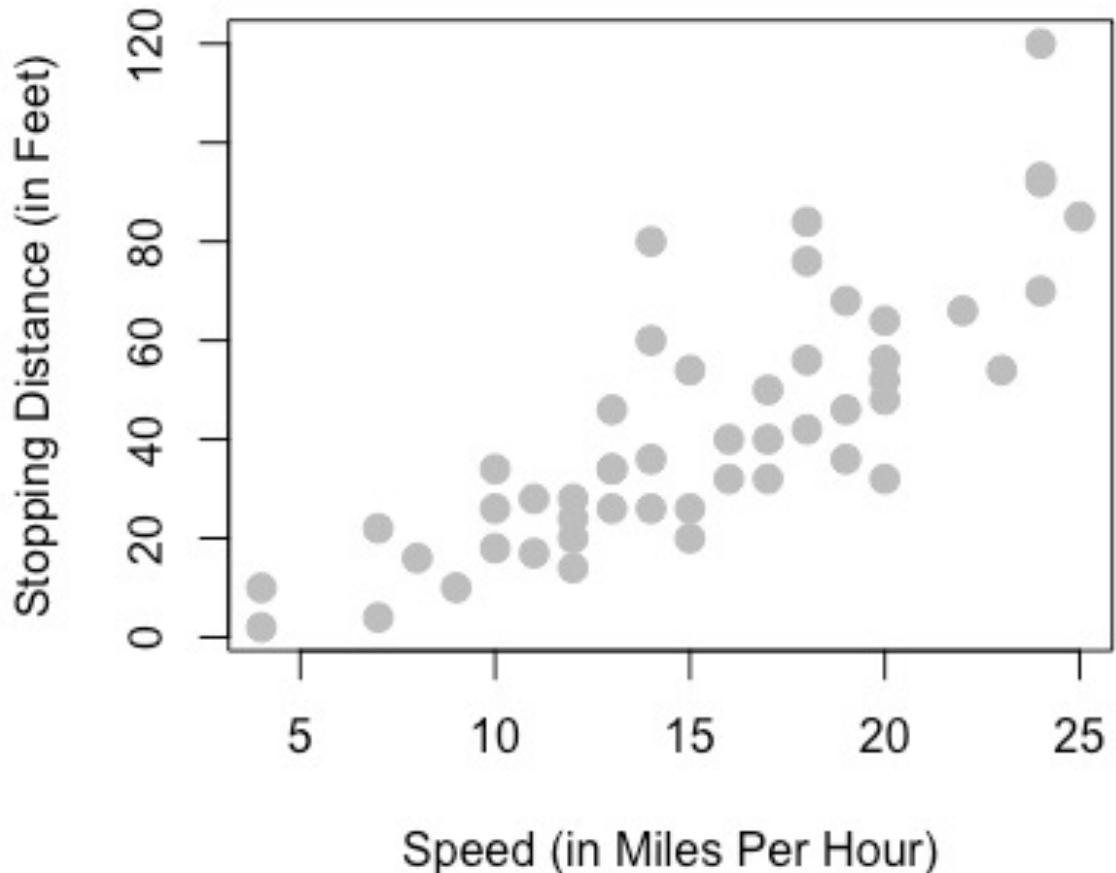
Stopping Distance vs Speed



- How would you describe the relationship between Stopping Distance and Speed?
 - Positive
 - Linear
 - Strong
 - Stopping distance is a function of speed

Hands On Example: Simple Linear Regression

Stopping Distance vs Speed



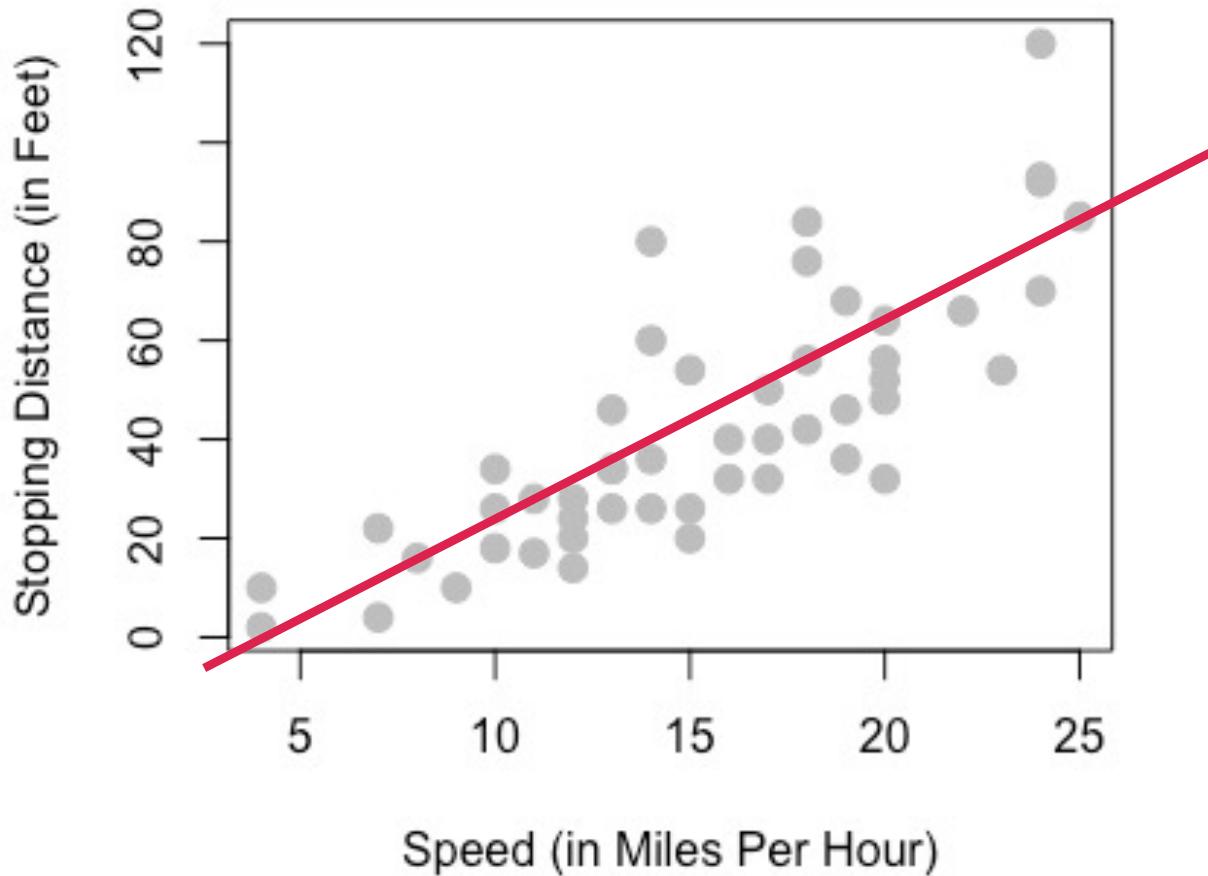
- How would you describe the relationship between Stopping Distance and Speed?
 - Positive
 - Linear
 - Strong
 - Stopping distance is a function of speed
- Could use a line to quantify the relationship between stopping distance and speed

$$y = \beta_0 + \beta_1 x$$

- y = stopping distance
- β_0 = y -intercept
- β_1 = slope
- x = speed

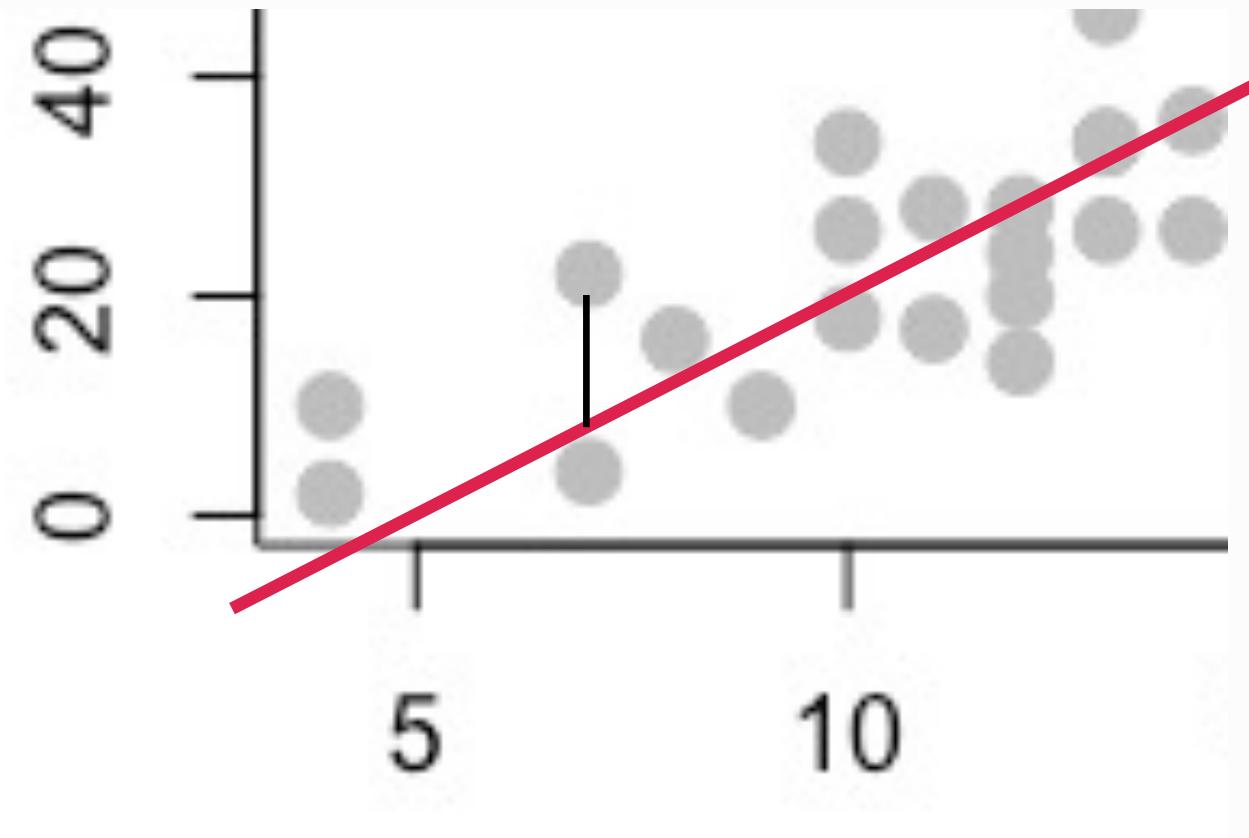
Hands On Example: Simple Linear Regression

Stopping Distance vs Speed



- How would you describe the relationship between Stopping Distance and Speed?
 - Positive
 - Linear
 - Strong
 - Stopping distance is a function of speed
- Could use a line to quantify the relationship between stopping distance and speed
$$y = \beta_0 + \beta_1 x$$
 - y = stopping distance
 - β_0 = y -intercept
 - β_1 = slope
 - x = speed
- How do we figure out the values of β_0 and β_1 ?

Hands On Example: Simple Linear Regression



- How do we figure out the values of β_0 and β_1 ?

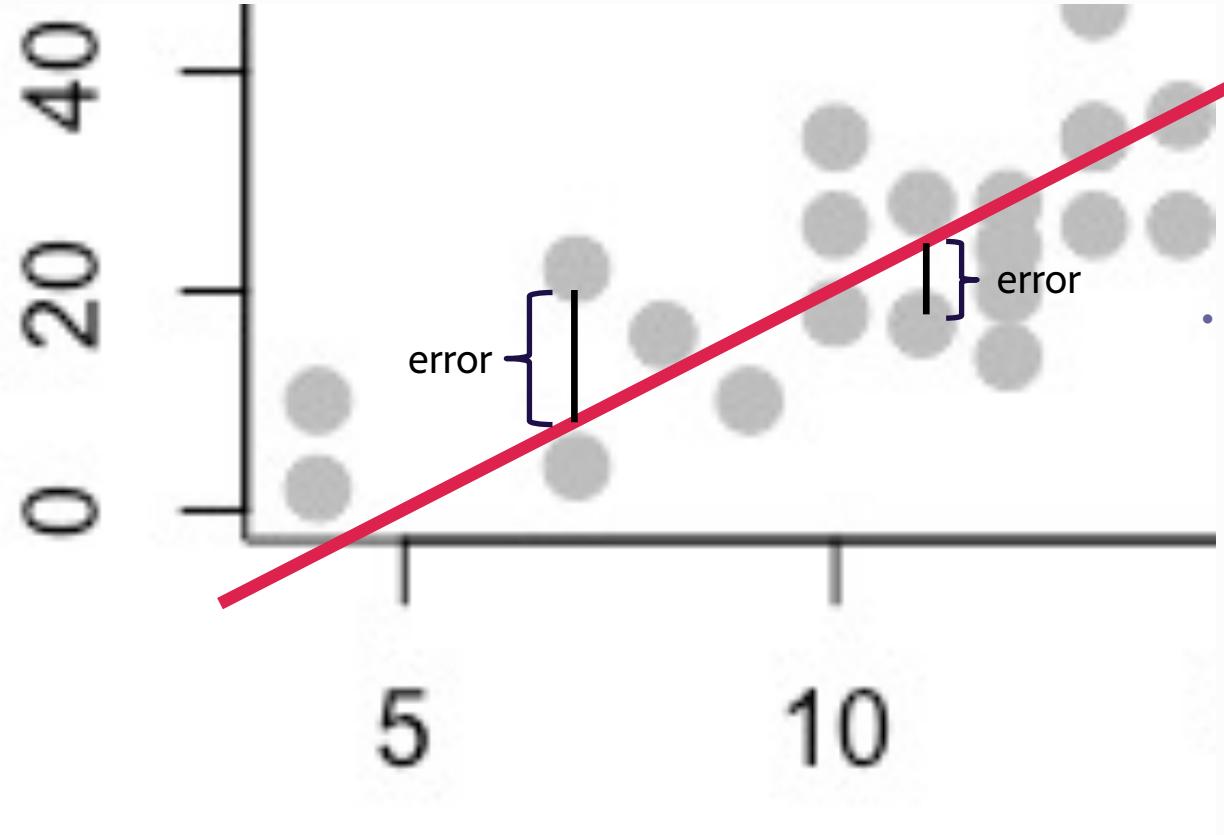
$$y = \beta_0 + \beta_1 x$$

- y = stopping distance
- β_0 = y-intercept
- β_1 = slope
- x = speed

Least-Squares Regression

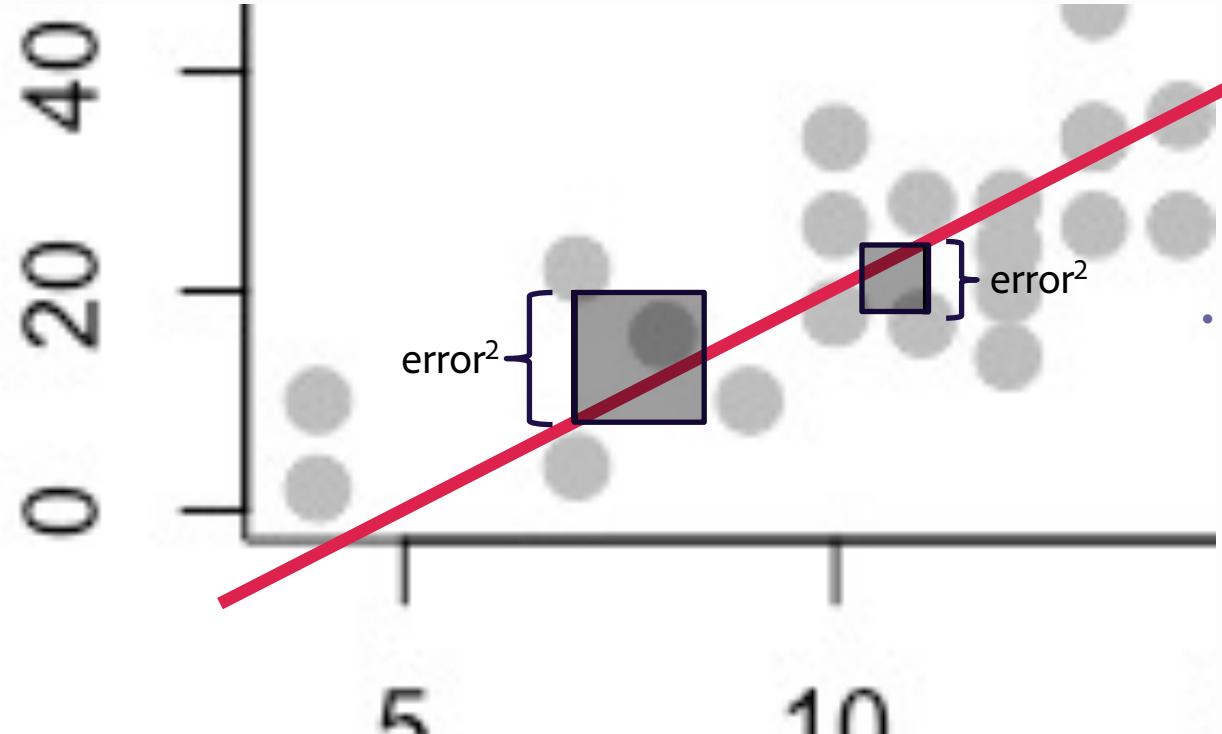
- Find values of β_0 and β_1 such that the sum of the squared error is minimized

Hands On Example: Simple Linear Regression



- How do we figure out the values of β_0 and β_1 ?
$$y = \beta_0 + \beta_1 x$$
 - y = stopping distance
 - β_0 = y-intercept
 - β_1 = slope
 - x = speed
- Least-Squares Regression
 - Find values of β_0 and β_1 such that the sum of the squared error is minimized

Hands On Example: Simple Linear Regression



- How do we figure out the values of β_0 and β_1 ?

$$y = \beta_0 + \beta_1 x$$

- y = stopping distance
- β_0 = y-intercept
- β_1 = slope
- x = speed

- Least-Squares Regression

- Find values of β_0 and β_1 such that the sum of the squared error is minimized

Hands On Example: Simple Linear Regression

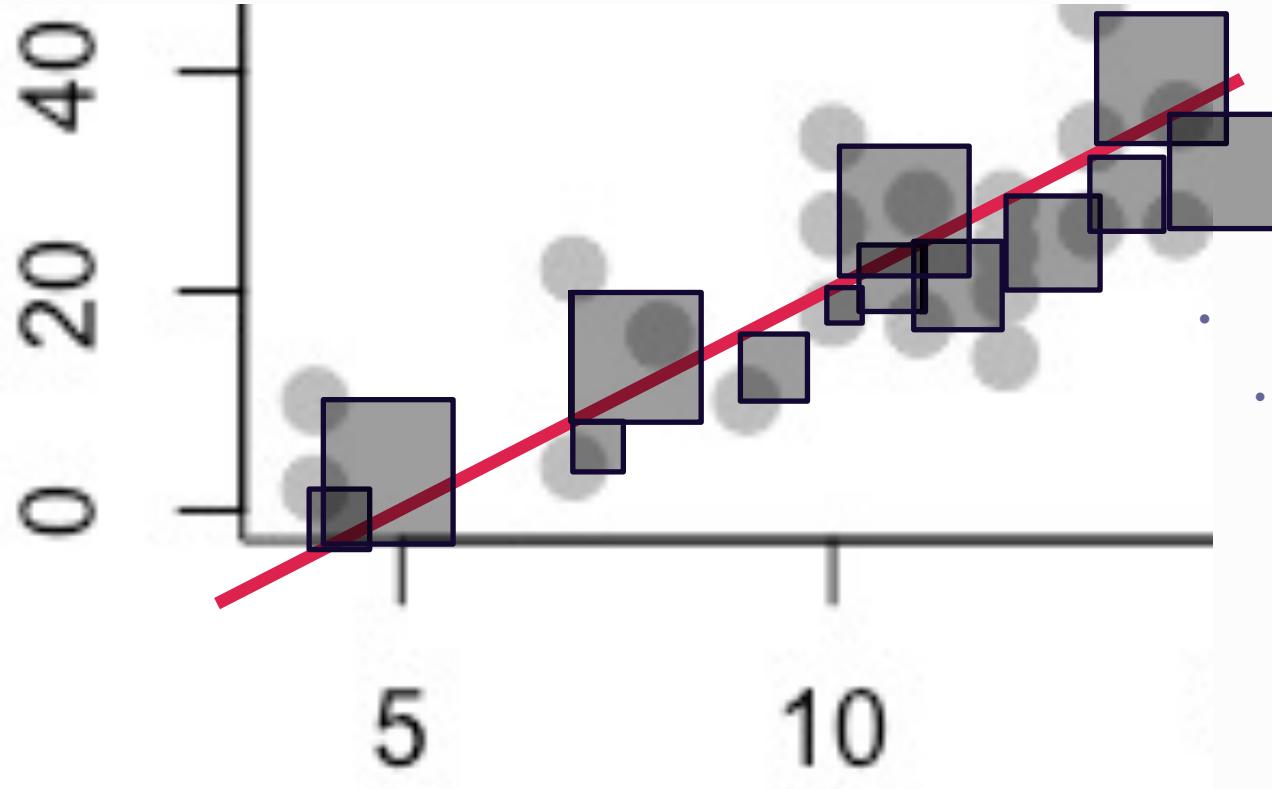
- How do we figure out the values of β_0 and β_1 ?

$$y = \beta_0 + \beta_1 x$$

- y = stopping distance
- β_0 = y-intercept
- β_1 = slope
- x = speed

- Least-Squares Regression

- Find values of β_0 and β_1 such that the sum of the squared error is minimized



Hands On Example: Simple Linear Regression

GITHUB REPO:

https://github.com/jlynlangford/tech_moms

FILE NAME:

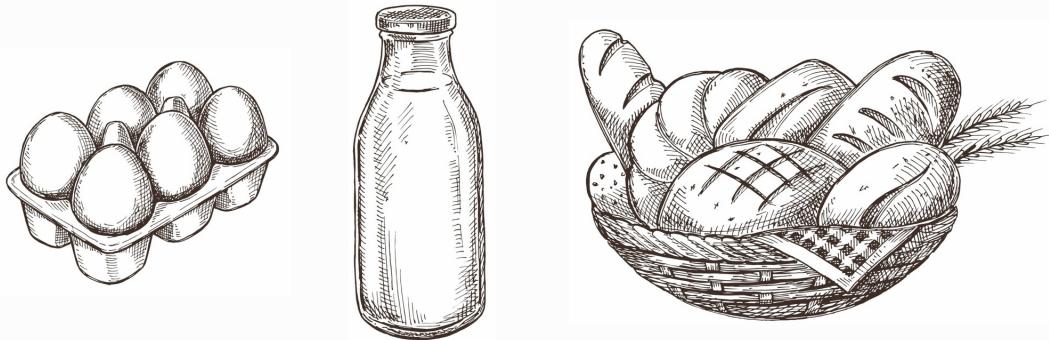
simple_linear_regression.R

Hands On Example: Product Recommendation Engine

The screenshot shows a product page for BEHR Premium Plus Interior Flat Paint. At the top left is a large image of a red paint can. To its right are four smaller images showing different sizes: 8 Ounce, 1 Quart, 1 Gallon, and 5 Gallon. Below these are sections for 'How to Get It' with options for 'Ship to Store' (free pickup), 'Ship to Home' (get it by Thursday, April 1, free with \$45 order), and 'Scheduled Delivery' (not available for this item). A 'REQUEST A PAINTER' button is also present. The price is listed as \$24.98. An 'Add to Cart' button is below the price. A red box highlights the 'Frequently Bought Together' section at the bottom, which lists three items: a paint can, a paint tray kit, and a paint brush set, all with a total price of \$50.52 and a 'Add all three to cart' button. A note states that shipping is unavailable to AK, GU, HI, PR, and VI.

- Product recommendation engines are used by many companies across many industries
- Algorithms and models can range from very simple to very complex
- Association Rules very simple way of finding connections in consumption behavior
 - Support – how often itemset $\{X, Y\}$ is found in the data set
 - Lift – expected Support if X and Y were independent
 - Confidence – proportion of transactions of X that also contain Y

Hands On Example: Product Recommendation Engine



TransactionID	Product Set
000001	{milk, eggs}
000002	{milk, eggs, bread}
000003	{milk, bread}
000004	{milk, bread}
000005	{milk, eggs, bread}
000006	{eggs}
000007	{eggs, bread}

Support {milk, eggs} = 3/7

Confidence {milk, eggs} = 3/5

$$\begin{aligned}\text{Lift } \{\text{milk, eggs}\} &= (3/5)/(3/7) \\ &= 21/15 \\ &= 1.4\end{aligned}$$

Lift > 1: Positive Effect (compliment)

Lift < 1: Negative Effect (substitute)

- Product recommendation engines are used by many companies across many industries
- Algorithms and models can range from very simple to very complex
- Association Rules very simple way of finding connections in consumption behavior
 - Support – how often itemset $\{X, Y\}$ is found in the data set
 - Confidence – proportion of transactions of X that also contain Y
 - Lift – expected Support if X and Y were independent

Hands On Example: Product Recommendation Engine

GITHUB REPO:

https://github.com/jlynlangford/tech_moms

FILE NAME:

recommender_dw.R

Q&A



MAKE IT AN EXPERIENCE