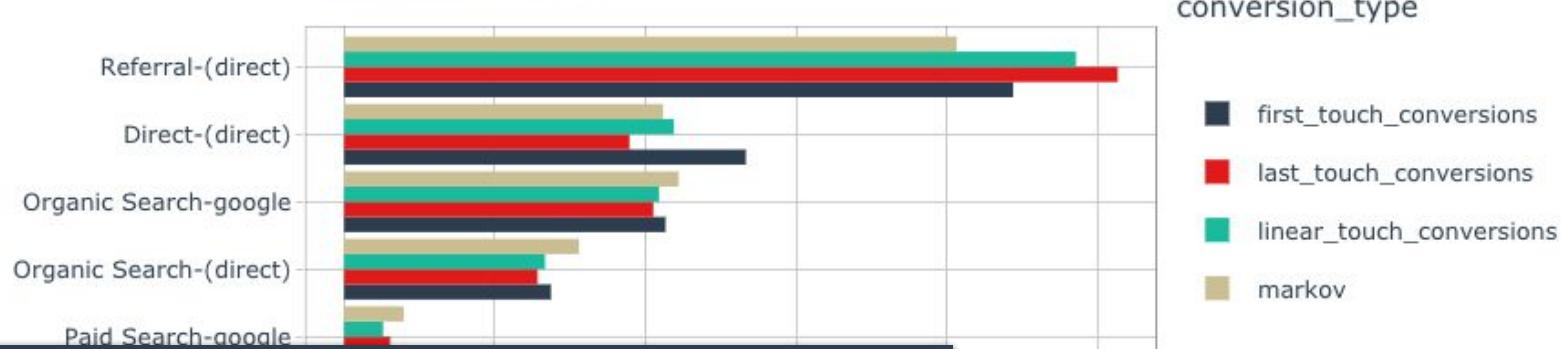


## Attribution Model



# Marketing Multi-Channel Attribution

## Using Google Analytics Data



5

Matt Dancho & David Curry  
*Business Science Learning Lab*





# Learning Lab Structure

- **Presentation**  
(20 min)
- **Demo's**  
(30 min)
- **Pro-Tips**  
(15 mins)



**Matt Dancho**

Founder of Business Science, Matt designs and executes educational courses and workshops that deliver immediate value to organizations. His passion is up-leveling future data scientists coming from untraditional backgrounds.



**David Curry**

Founder of Sure Optimize, David works with businesses to help improve website performance and SEO using data science. His passion is ethical Machine Learning initiatives.

# Success Story

## Chris Selig

- Data Consultant
- Took Shiny Dev Course
- Made Resume in ...

**SHINY!!!**



#Business  
Science  
Success

*"Thank you Matt for the R Shiny Course at Business Science."*



Shiny Resume for Chris Selig

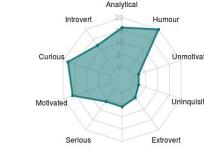
### About Me

Broke into the data ninja field through my business intelligence (BI) work with extensive experience gathering requirements and building automated reports and dashboards for multiple industries, including education, insurance, airport security and oil and gas and many different departments, including information technology, finance, vendor management, procurement, health/safety and environment. Loved that work but wanted to do more so now I have transitioned into advanced analytics and data science where my true passions lie. My greatest strength is being a strong communicator and can speak "IT" as well as "business" which is extremely helpful when I need to communicate highly technical material to non-technical audiences.

### Experience

BIDAMIA INC., Calgary, AB 2018 - Present  
Business Intelligence & Advanced Analytics Consultant  
Development of both Spotfire and Power BI dashboards/reports for clients, provided training to end users, and developed ETL processes using KNIME  
Client: Husky Energy  
Using R, completed a text mining analysis of software vendor responses to aid in the selection of an Information Technology Financial Management (ITFM) tool.  
Using R, developed a machine learning model (using XGBoost algorithm) to predict monthly Information Services (IS) expenses  
Developed a time series forecasting model using seasonal naive algorithm to quickly forecast Information Service expenses.  
Developed a show back report, using R and Spotfire, that for the first time, enabled the organization for the first time to delve deep into IS IS costs for the business unit and to begin rationalizing the environment  
Created an automated ETL process using KNIME that combined forecast/budget files and loaded to a SQL Server database.  
Using SQL Server, developed a database as a single source of truth for Planning and Controls team. Database provides significant reduction in reporting time as endless spreadsheets no longer need to be compiled for reporting requests from management.  
Developed a dashboard for Power BI to various groups within Canada and the US  
Created a Power BI report using R and Python to generate a language to develop custom visualizations within Power BI

### Personality



### Motivators

Cool colleagues that can laugh at/me  
Environments where I learn everyday  
Complex problems to solve

### Top Skills

Data Visualization

### Education

Certificate in Analytics for Data Science, 2020, Institute for Statistics Education  
Certificate in Business Intelligence and Analytics, University of Calgary, AB  
Masters of Business Administration (MBA), Finance, Dalhousie University, Halifax, NS  
Bachelor of Arts (BA), Economics, Acadia University, Wolfville, NS

### Professional Development

<https://chris-selig.shinyapps.io/ShinyResume/>

# Marketing Series



**Learning Labs Pro**  
Community-Driven Data Science Courses  
 Matt Dancho 

- **Lab 24 - A/B Testing**
  - Business Science's Website
  - Infer - Bootstrap & Permutation
- **Lab 25 - Multi-Channel Attribution (Part 1)**
  - Google Analytics Data
  - ChannelAttribution
- **Lab 26 - Multi-Channel Attribution (Part 2)**
  - Multi-Touch Channel Attribution
  - Costs, Path Splitting, Network Visualization
- **Lab 27 - Automated Prediction & Tracking Google Trends**
  - Google Trend Forecasting
  - gtrendsR, forecast
  - chronR, taskscheduleR



# Learning Labs PRO

Every 2-Weeks

1-Hour Course

Recordings + Code + Slack

**\$19/month**

*university.business-science.io*

*Lab 24*  
**A/B Testing with Infer**

*Lab 23*  
**SQL with BigQuery & Conversion Funnel**

*Lab 22*  
**SQL for Time Series**

*Lab 21*  
**SQL for Data Science**

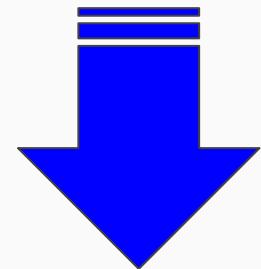
*Lab 20*  
**Explainable Machine Learning**

*Lab 19*  
**Using Customer Credit Card History for Networks Analysis**

*Lab 18*  
**Time Series Anomaly Detection with anomalize**



**Continuous Learning**  
Advanced Topics



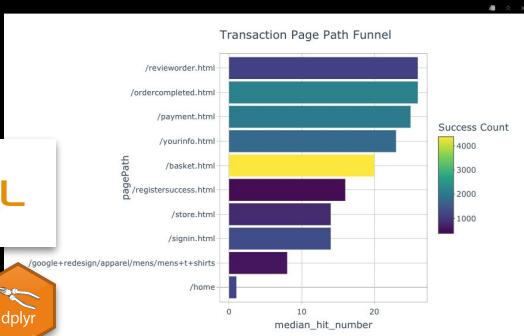
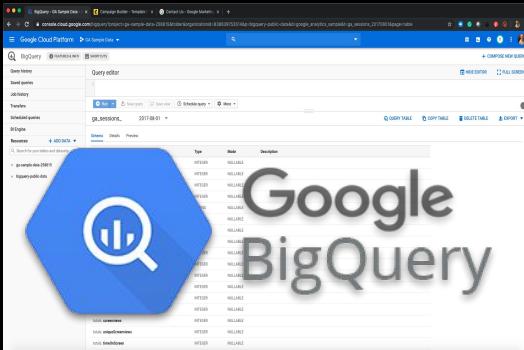
**Learning Labs Pro**

Community-Driven Data Science Courses

 Matt Dancho

**\$19/m**

# Agenda



- **Business Case Study**
  - Google Merchandise Store
  - Channel Attribution
- **30-Min Demo**
  - Multi-Touch Channel Attribution
- **Google Analytics**
  - Terminology
- **Pro-Tips & Learning Guide**
  - Recap + Pro-Tips
  - Learning Plan
- **Big Query**
  - 80/20 Data Concepts



# Google Merchandise Store

Business Case



# Google Merchandise Store

## Google Analytics Data

**Customers** can purchase t-shirts, gear, etc

**Google Analytics** tracks every event on the website.

We can use this for **Channel Attribution**.

<https://shop.googlemerchandise.com/>

The screenshot shows a web browser displaying the Google Merchandise Store at [shop.googlemerchandisestore.com](https://shop.googlemerchandisestore.com). The page features a header with the Google logo and 'official merchandise store'. Below the header is a large image of several Google t-shirts in red, green, and blue, with the text 'These tees are the bee's knees' and 'The speckled Google T-shirt is just one of the new tees this season.' A prominent blue button labeled 'Shop now' is visible. At the bottom, there are contact details: a phone icon with '1-855-300-2945', an envelope icon with 'Email Us', and a 'Customer Support' link. The footer also includes the text 'Then go outside. We will take it.'



# Channel Attribution

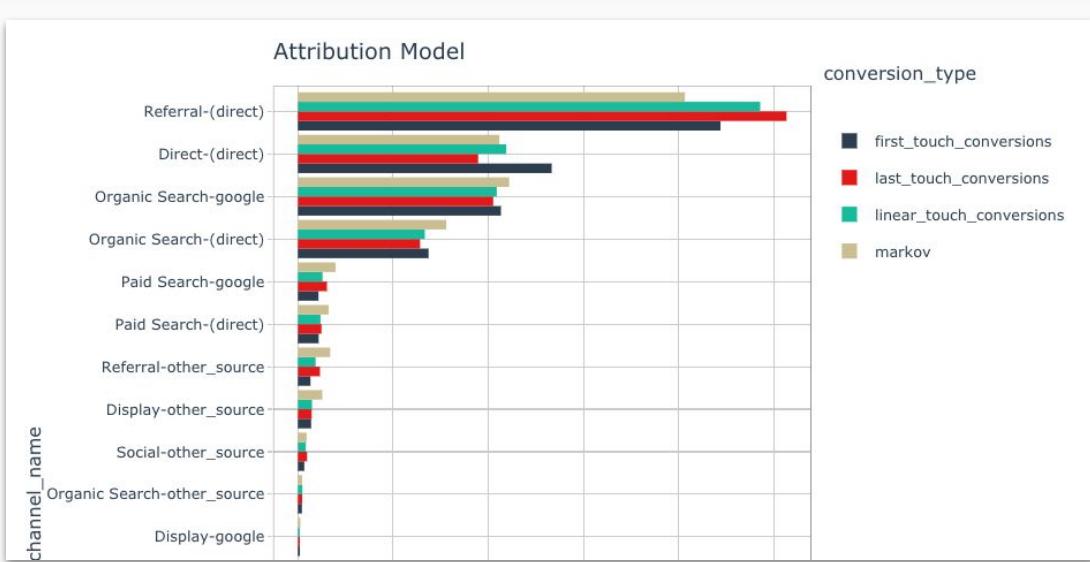
## Conversion Models

Many Types

Last Touch (most common)

### Problem:

Doesn't take into account supporting channels





# Marketing Budget

## Allocate Time & Spend

If you use Last Touch, your supporting channels are underfunded, hurting your conversion funnel.

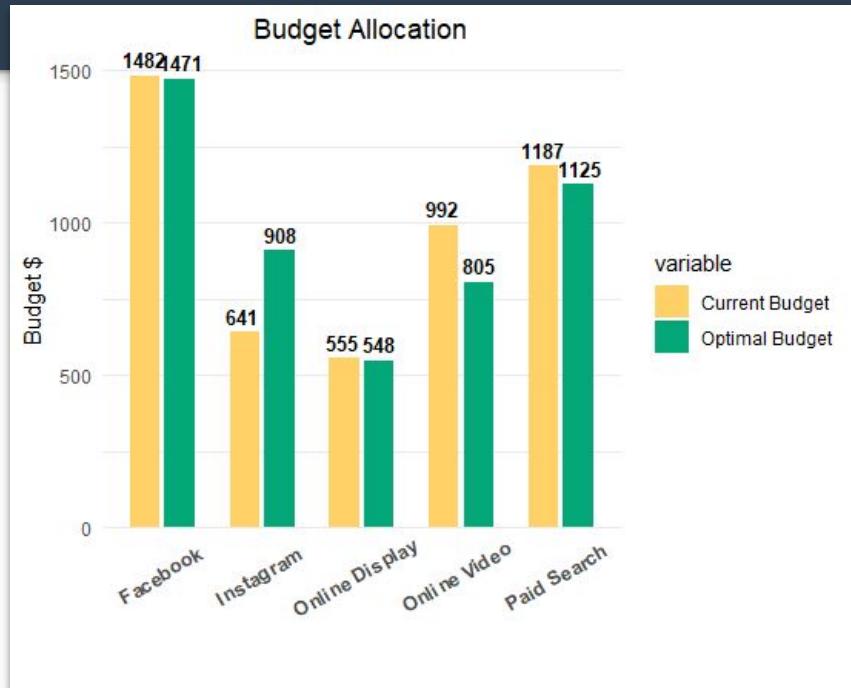


Image Credit: <https://github.com/MatCyt/Markov-Chain>

# Customer Journey Concepts

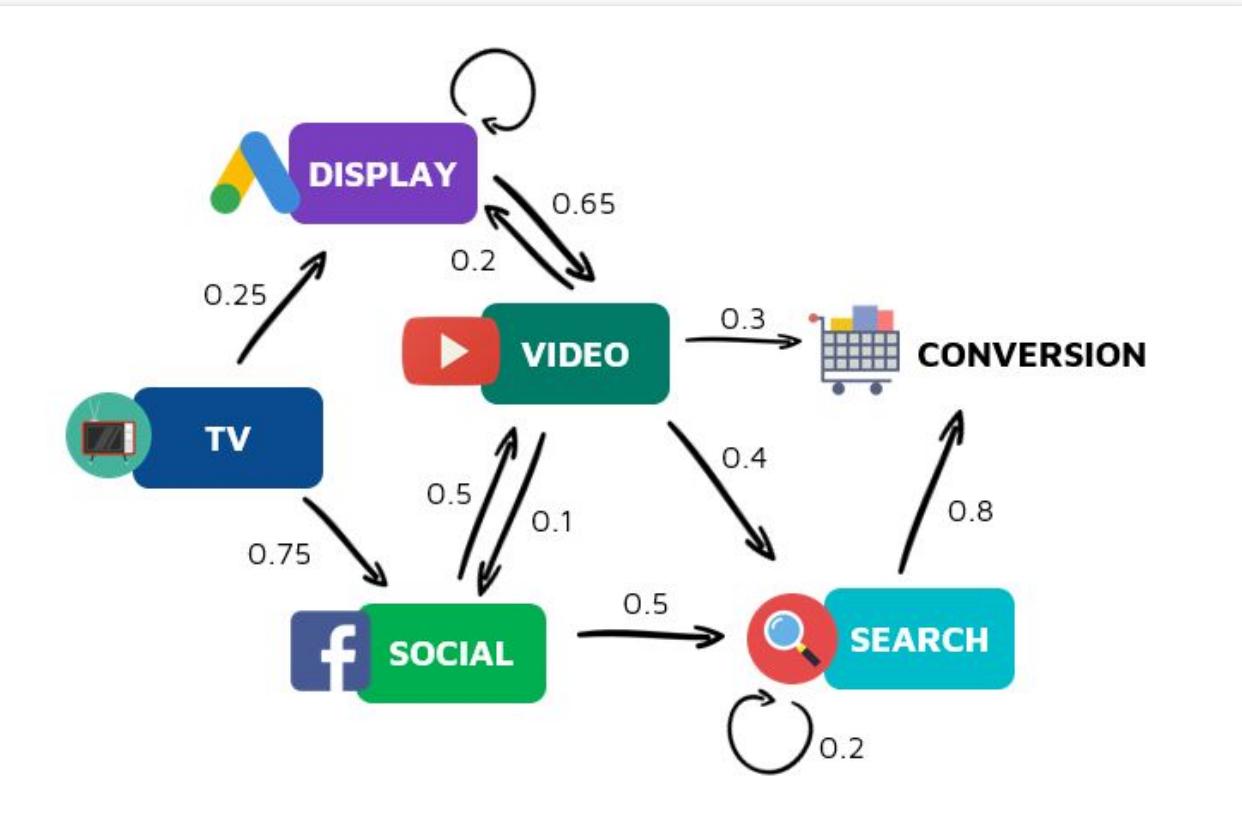


## Touch Points

User interacts with media, website, referral, social, and search.

Tracked in **Google Analytics**:

- User ID, Session ID
- Channel Group
- Traffic Source



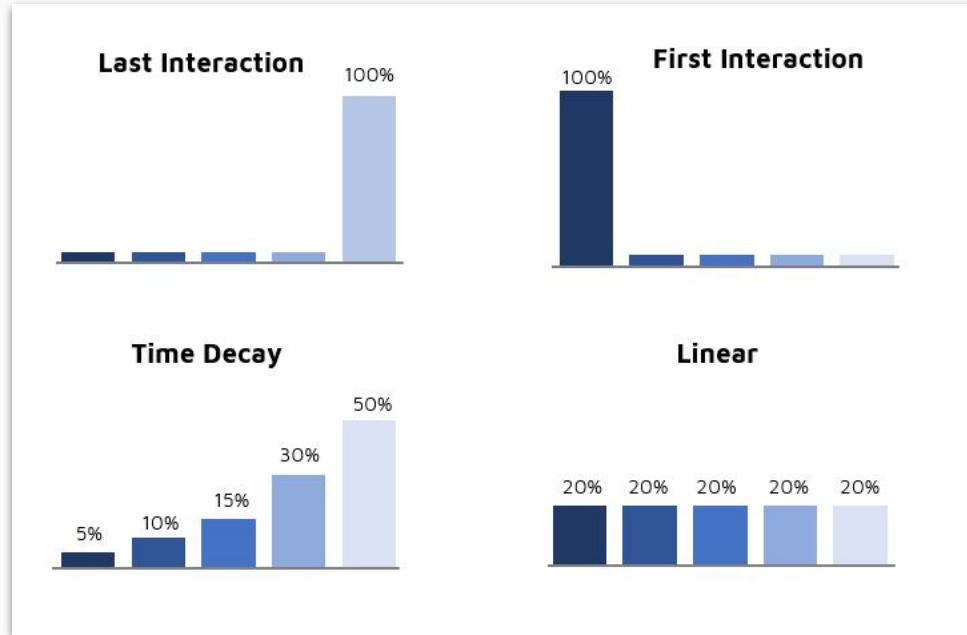


## Heuristic Models

Many methods exist, but these **fail to account for supporting touch points.**

**Pros:** Simple

**Cons:** Incorrect Account for Supporting Touch Points





# Better Method?

What if we use Game Theory?

## Markov Chains

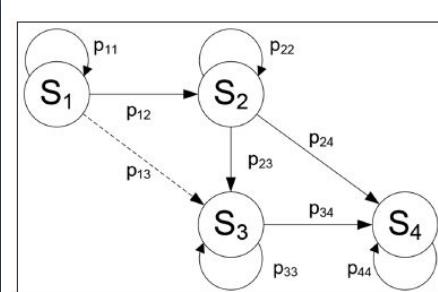


Fig. 1: Transition diagram for the Markov Chain. S (states) represents four different stages of the progression of the disease among vulnerable, HIV infective, clinical AIDS persons, and death. The transient states are S1, S2 and S3, while the recurrent absorbing state is S4. Transitions between these states will be modelled separately for the African Americans and Caucasians.

# Markov Chains

Most popular method for attribution modeling

# Markov Chains

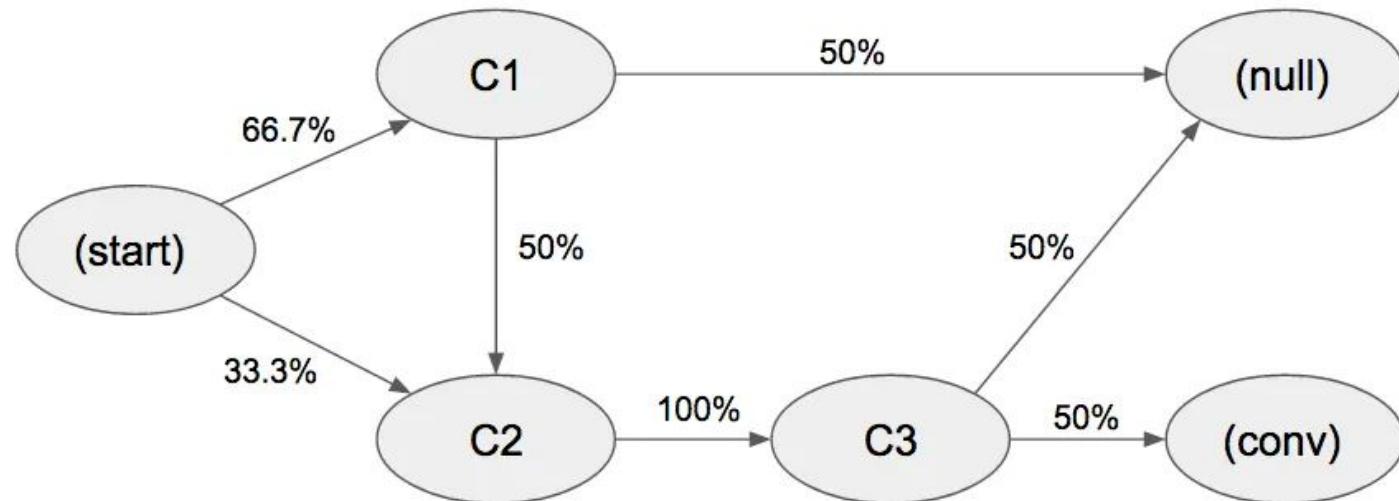


Image Source: <https://analyzecore.com/2016/08/03/attribution-model-r-part-1/>

# Markov Chains

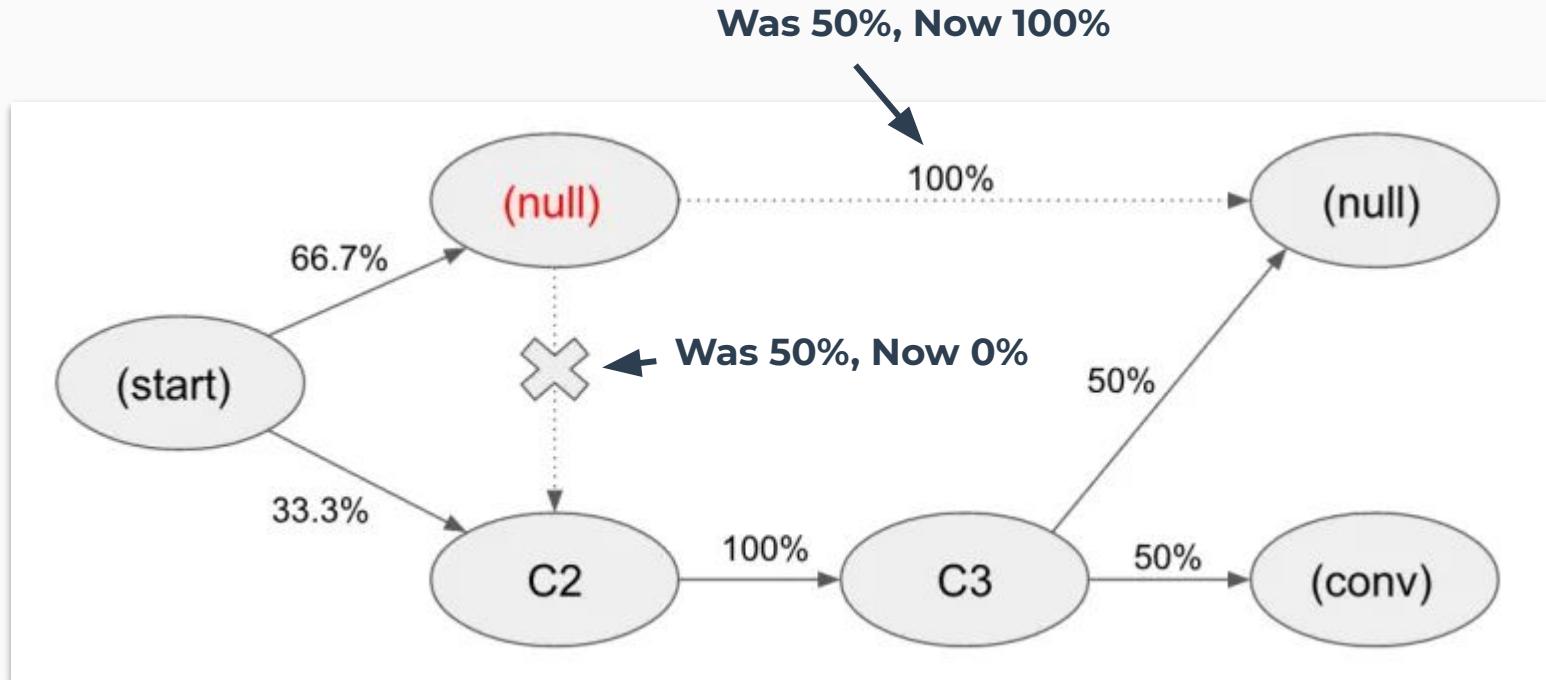


Image Source: <https://analyzecore.com/2016/08/03/attribution-model-r-part-1/>

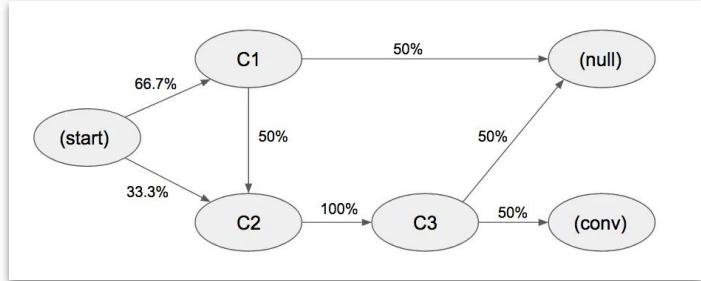


## Calculate Removal Effect

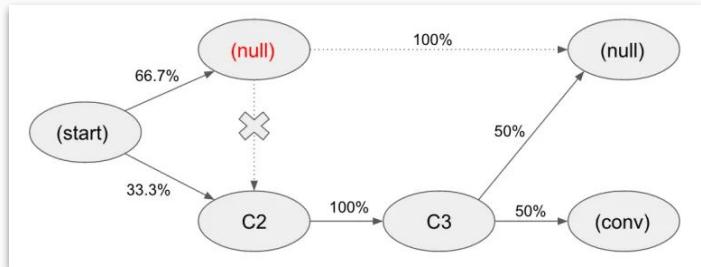
Complete Model: **33%**

C1 Removed: **16.7%**

Removal Effect: **50%**

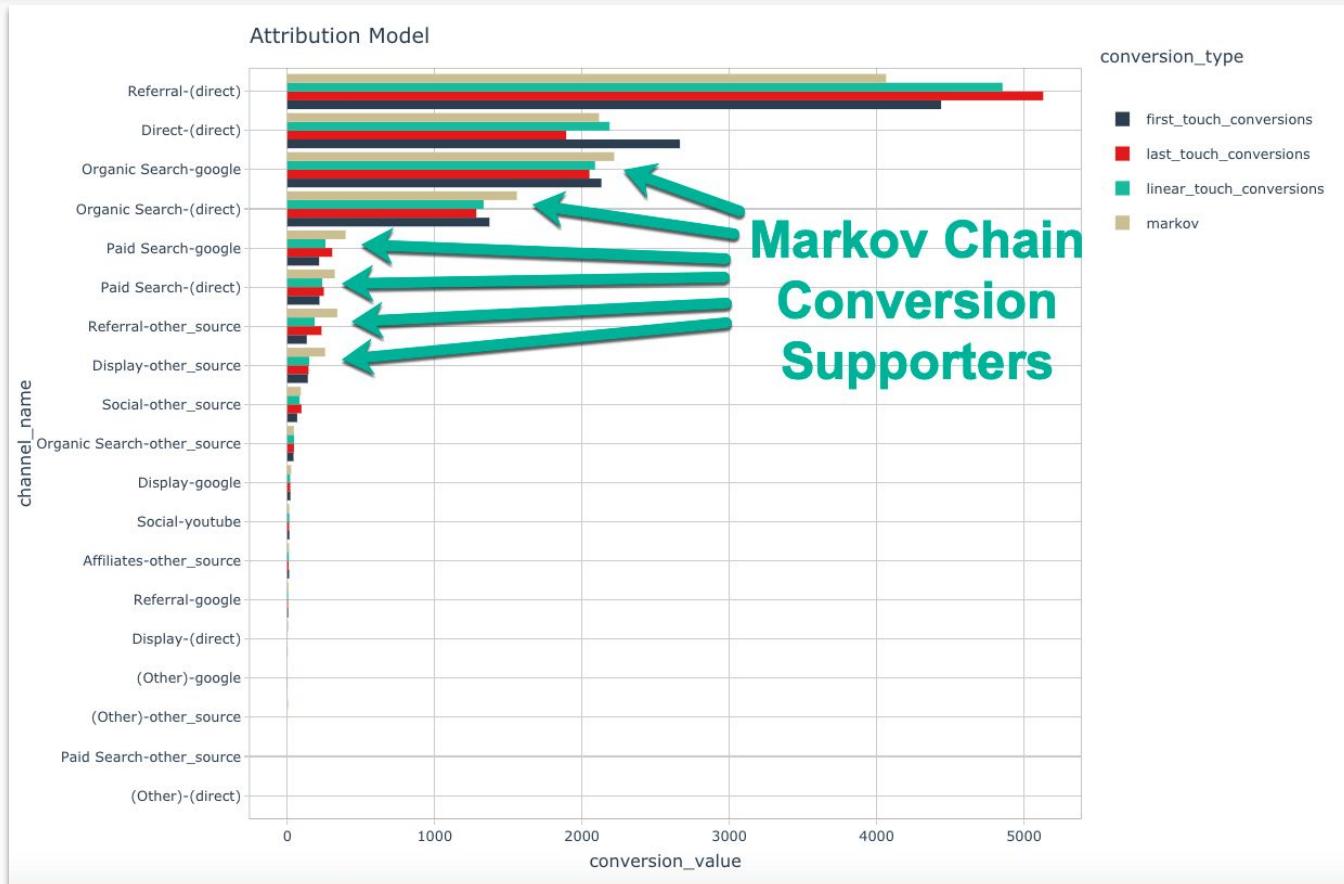


$$P_{\text{conv}} = 0.667 * 0.5 * 1 * 0.5 + 0.333 * 1 * 0.5 = 33.3\%$$



$$P_{\text{conv}} = 0.333 * 1 * 0.5 = 16.7\%$$

# Markov Chains



# 30-Min Demo

Channel Attribution w/  
Google Analytics Data

# PRO-TIPS

Yeahhhhhh!

# Pro-Tips #1 - Split Channel Paths



## We did not split paths

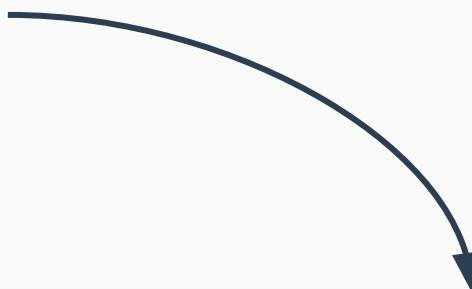
Will show how in Part 2 - Modifying this code

### We treated this path as 1 path:

- C1 > C2 > (conversion) > C1 > (conversion)

### Should be split into:

- Path 1: C1 > C2 > (conversion)
- Path 2: C1 > (conversion)



```
246 # 3.2 CHANNEL-PATH MANIPULATION ----
247 channel_path_dtplyr <- visitor_date_channelgrouping_dtplyr %>%
248
249     group_by(fullVisitorId) %>%
250     summarize(
251         channel_path      = str_c(channel_source, collapse = " > "),
252         conversion_total = sum(total_transactions),
253         conversion_null = sum(total_transactions == 0),
254         conversion_value = sum(total_transaction_revenue),
255         n_channel_path   = n()
256     ) %>%
257     ungroup()
258
259 channel_path_tbl <- channel_path_dtplyr %>% as_tibble()
260
```

# Pro-Tips #2 - Assign Cost Weights

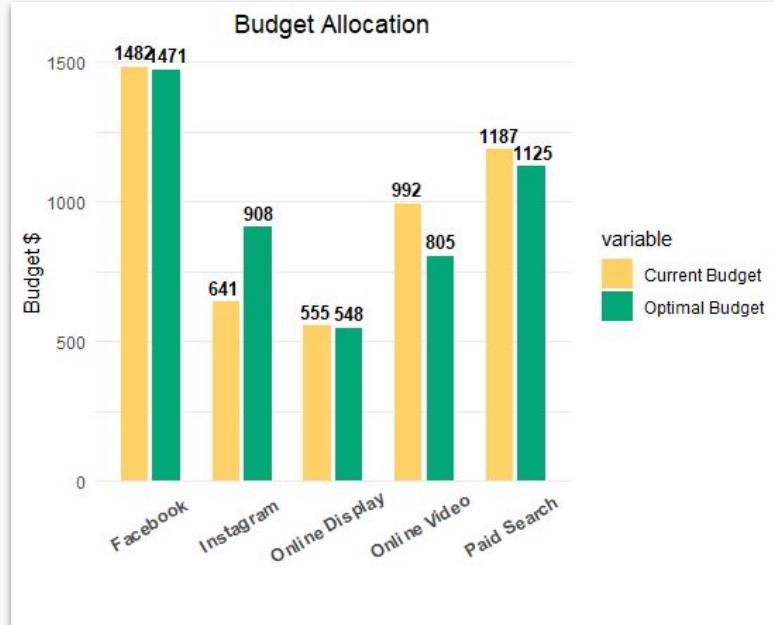


## We did not include costs

Will show how in Part 2

### Calculate Return on Ad Spend (ROAS)

- ROAS = Channel Conversion Weight / Channel Budget Weight



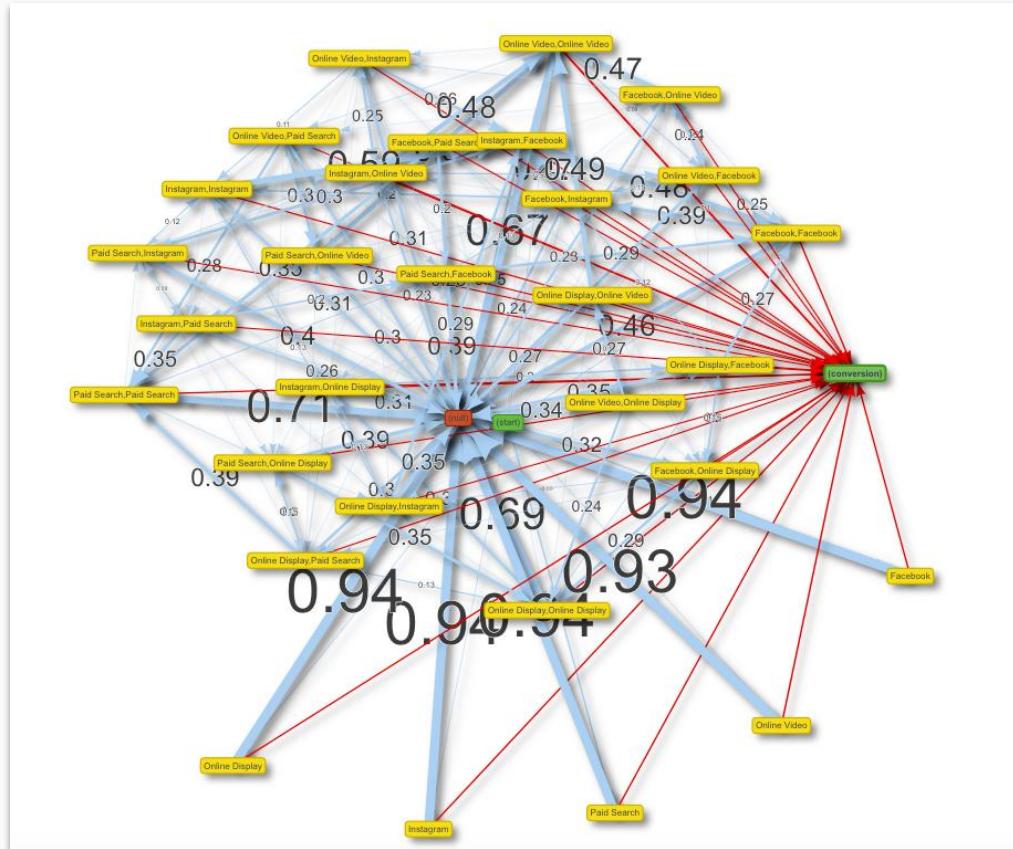
# Pro-Tips #3 - Visualize the Channel Network



## We did not visualize

Useful for understanding which combinations result in increased conversion.

Will show how in Part 2



# Pro-Tips #4 - Machine Learning



## Alternative Methods

Use ML for prediction, which goes beyond probability to linear and tree-based methods.

- Random Forest
- GLM
- XGBoost

### Data-driven Multi-touch Attribution Models

Xuhui Shao<sup>\*</sup>  
Turn, Inc.  
835 Main St.  
Redwood City, CA 94063  
xuhui.shao@turn.com

Lexin Li<sup>†</sup>  
Department of Statistics  
North Carolina State University  
Raleigh, NC 27695  
li@stat.ncsu.edu

#### ABSTRACT

In digital advertising, attribution is the problem of assigning credit to one or more advertisements for driving the user to the desirable actions such as making a purchase. Rather than giving all the credit to the last ad, accurate multi-touch attribution allows more than one ads to get the credit based on their corresponding contributions. Multi-touch attribution is one of the most important problems in digital advertising, especially when multiple media channels, such as search, display, mobile and social are involved. Due to the lack of statistical framework and available modeling approach, true data-driven methodology does not exist today in the industry. While predictive modeling has been thoroughly researched in recent years in the digital advertising domain, the attribution problem focuses more on accurate and stable interpretation of the influence of each user interaction to the final user decision rather than just user classification. Traditional classification models fail to achieve those goals.

In this paper, we first propose a bivariate metric, one measures the variability of the estimate, and the other measures the accuracy of classifying the positive and negative users. We then develop a bagged logistic regression model, which we show achieves a comparable classification accuracy as a usual logistic regression, and a more stable estimate of individual advertising channel contribution. We also propose an intuitive and simple probabilistic model to directly quantify the attribution of different advertising channels. We then apply both the bagged logistic model and the probabilistic model to a real-world data set from a multi-channel advertising campaign for a well-known consumer software and service brand. The two models produce consistent general conclusions and thus offer useful cross-validation. The

<sup>\*</sup>Xuhui Shao is Chief Technology Officer, Turn, Inc.

<sup>†</sup>Lexin Li is the corresponding author and Associate Professor, Department of Statistics, North Carolina State University.

results of our attribution models also shed several important insights that have been validated by our advertising team. We believe applying data-driven predictive models in the production advertising platform of the first author's company, and plan to implement the bagged logistic regression in the next product release. We believe availability of such data-driven multi-touch attribution metric and models is a breakthrough in the digital advertising industry.

#### Categories and Subject Descriptors

I.6.5 [Computing Methodologies]: Simulation and Modeling, Model Development

#### General Terms

Algorithms, Performance, Theory

#### Keywords

Digital Advertising, Multi-touch Attribution Model, Bagged Logistic Regression

#### 1. INTRODUCTION

Digital advertising started 16 years ago as a new media where traditional print ads can appear [1]. When internet continues to grow with an exploding rate, advertising industry embraced digital advertising and has made it a \$40 Billion a year mega industry in US alone. Digital advertising's appeal not only in its ability to precisely target different groups of consumers with customized ad messages and ad placements, but also its great improvements in its ability to track revenue and performances almost instantaneously.

Advertising campaigns are often launched across multiple channels. Traditional advertising channels include outdoor billboard, TV, radio, newspapers and magazines, and direct mailing. Digital advertising channels include search, online display, social, video, mobile and email. In this article, we focus on the digital advertising channels. Typically, multiple advertising channels have delivered advertisement impressions to a user. When the user then makes a purchase decision or signs up to a service being advertised, the advertiser wants to determine which ads have contributed to the user's decision. This step is critical in completing the feedback loop so that one can analyze, report and optimize an advertising campaign. This problem of interpreting the influence of advertisements to the user's decision process is called the attribution problem.

# Marketing Series



**Learning Labs Pro**  
Community-Driven Data Science Courses  
 Matt Dancho 

- **Lab 24 - A/B Testing**
  - Business Science's Website
  - Infer - Bootstrap & Permutation
- **Lab 25 - Multi-Channel Attribution (Part 1)**
  - Google Analytics Data
  - ChannelAttribution
- **Lab 26 - Multi-Channel Attribution (Part 2)**
  - Costs, Path Splitting, & Network Visualization
  - Machine Learning
- **Lab 27 - Automated Prediction & Tracking Google Trends**
  - Google Trend Forecasting
  - gtrendsR, forecast
  - chronR, taskscheduleR



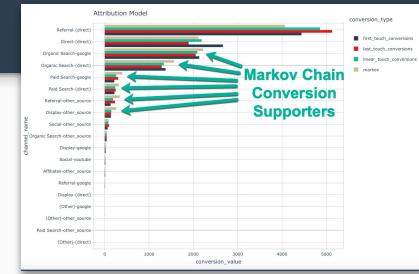
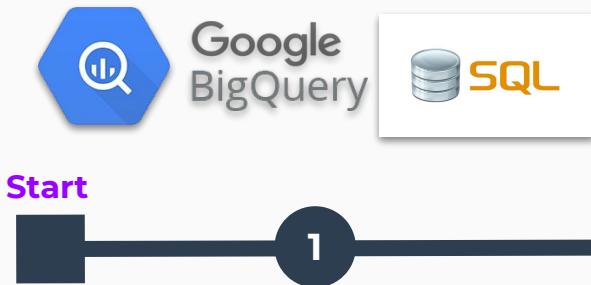
# **What We Just Did**

And how WE did it!



# Channel Attribution Workflow

## Step-By-Step



### BigQuery

Connect to BigQuery database  
containing Google Analytics sessions  
data

### dplyr, dtplyr & ggplot2

Big Data Wrangling, Aggregation, &  
Visualization

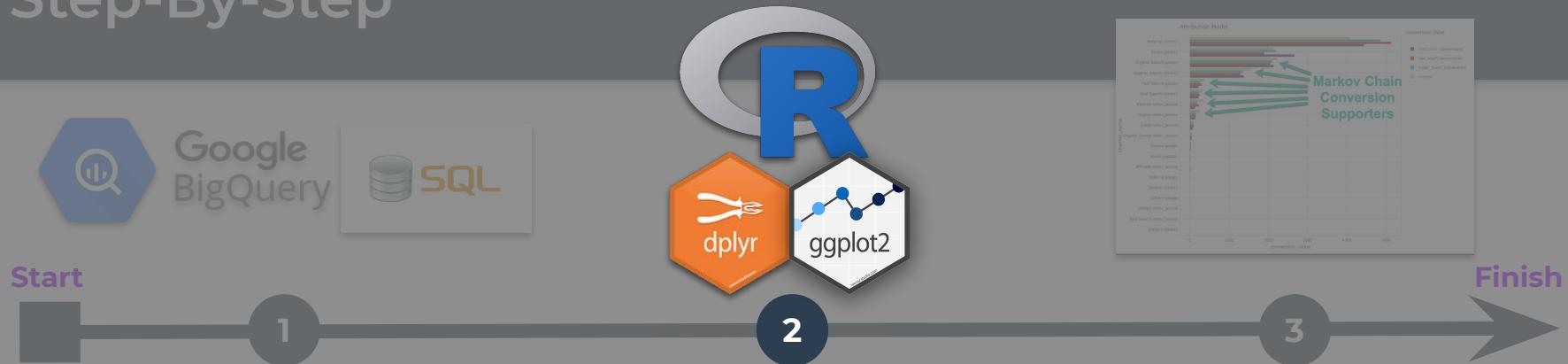
### ChannelAttribution

Visualize Conversion Funnel  
Detect Most Important Step(s)

**What is the most important step?**

# Channel Attribution Workflow

## Step-By-Step



### BigQuery

Connect to BigQuery database  
containing Google Analytics sessions  
data

### dplyr, dtplyr & ggplot2

Big Data Wrangling, Aggregation, &  
Visualization

### Channel Attribution

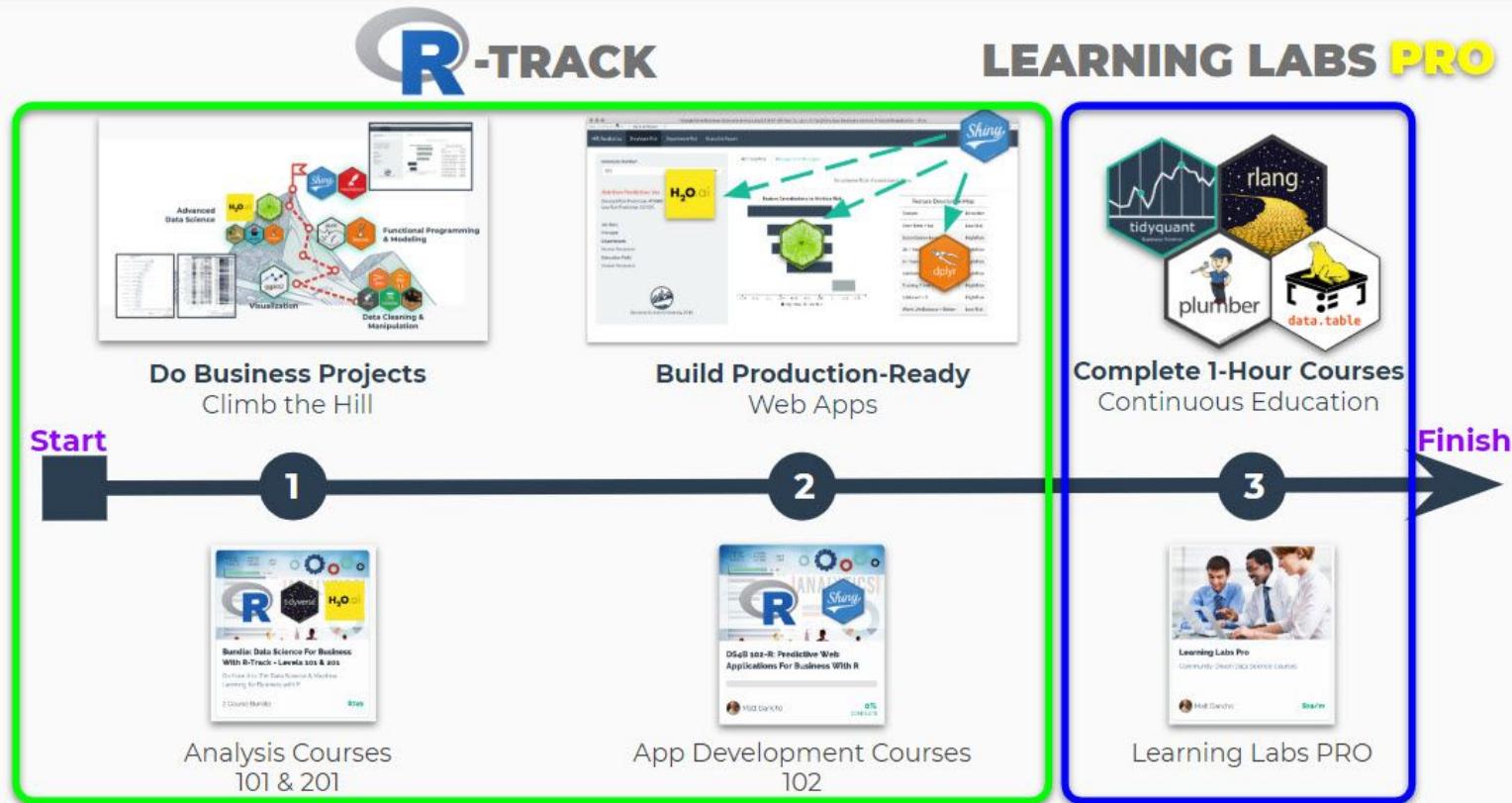
Visualize Conversion Funnel  
Detect Most Important Step(s)

**What is the most important step?**

# **Business Science University**

Learn Data Science for Business with R in **6-Months**

# The program that will deliver YOUR Transformation



Everything is **Taken Care of** For You in Our Platform

# 4-Course R-Track System



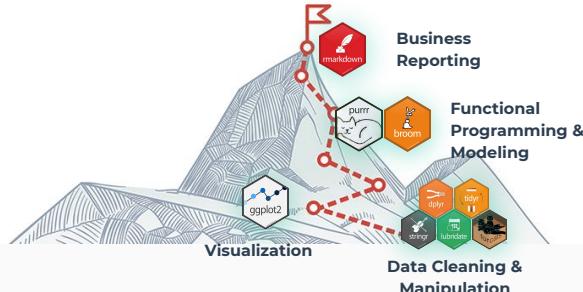
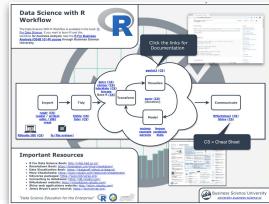
## Business Analysis with R (DS4B 101-R)

## Data Science For Business with R (DS4B 201-R)

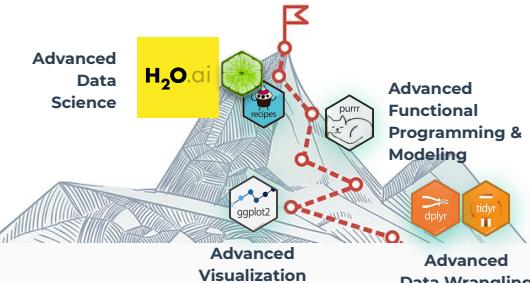
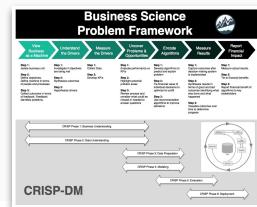
## Web Apps & Shiny Developer (DS4B 102-R + DS4B 202A-R)

### Project-Based Courses with Business Application

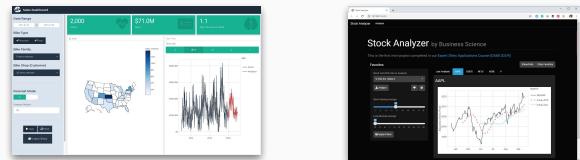
Data Science Foundations  
**7 Weeks**



Machine Learning & Business Consulting  
**10 Weeks**



Web Application Development  
**12 Weeks**

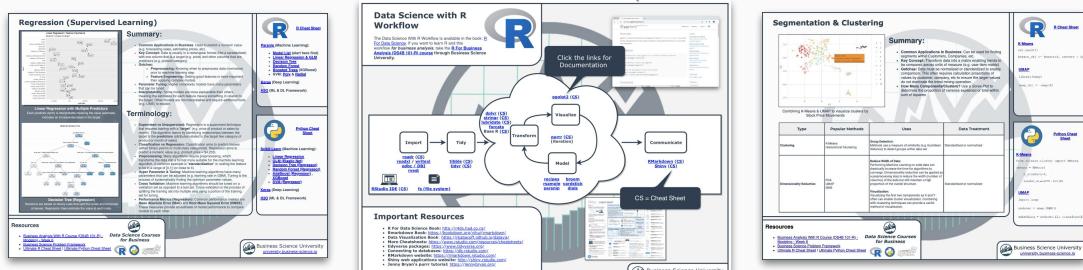


# Key Benefits

- Fundamentals - Weeks 1-5 (25 hours of Video Lessons)
  - Data Manipulation (dplyr)
  - Time series (lubridate)
  - Text (stringr)
  - Categorical (forcats)
  - Visualization (ggplot2)
  - Programming & Iteration (purrr)
  - 3 Challenges
- **Machine Learning - Week 6 (8 hours of Video Lessons)**
  - Clustering (3 hours)
  - Regression (5 hours)
  - 2 Challenges
- Learn Business Reporting - Week 7
  - RMarkdown & plotly
  - 2 Project Reports:
    1. Product Pricing Algo
    2. Customer Segmentation

# Business Analysis with R (DS4B 101-R)

Data Science Foundations  
**7 Weeks**



# Key Benefits

## End-to-End Churn Project

Understanding the Problem & Preparing Data - Weeks 1-4

- Project Setup & Framework
- Business Understanding / Sizing Problem
- Tidy Evaluation - rlang
- EDA - Exploring Data -GGally, skimr
- Data Preparation - recipes
- Correlation Analysis
- 3 Challenges

## Machine Learning - Weeks 5, 6, 7

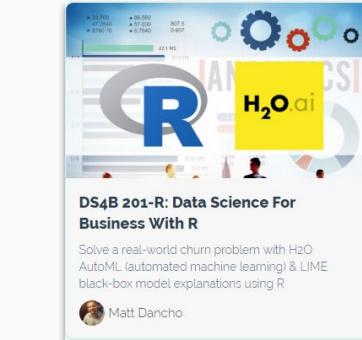
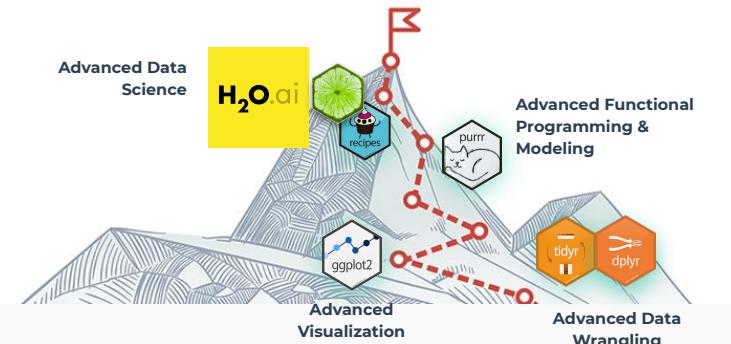
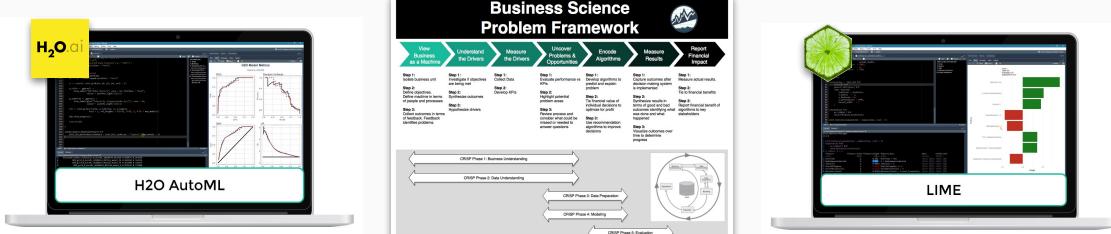
- H2O AutoML - Modeling Churn
- ML Performance
- LIME Feature Explanation

## Return-On-Investment - Weeks 7, 8, 9

- Expected Value Framework
- Threshold Optimization
- Sensitivity Analysis
- Recommendation Algorithm

# Data Science For Business (DS4B 201-R)

Machine Learning & Business Consulting  
**10 Weeks**



# Key Benefits

## Learn Shiny & Flexdashboard

- Build Applications
- Learn Reactive Programming
- Integrate Machine Learning

## App #1: Predictive Pricing App

- Model Product Portfolio
- XGBoost Pricing Prediction
- Generate new products instantly

## App #2: Sales Dashboard with Demand Forecasting

- Model Demand History
- Segment Forecasts by Product & Customer
- XGBoost Time Series Forecast
- Generate new forecasts instantly

# Shiny Apps for Business (DS4B 102-R)



Web Application Development  
**4 Weeks**



DS4B 102-R: Shiny Web Applications  
For Business (Level 1)

Build a predictive web application using Shiny, Flexdashboard, and XGBoost.

Matt Dancho

# Key Benefits

Frontend + Backend + Production Deployment

## Frontend for Shiny

- Bootstrap

## Backend for Shiny

- MongoDB
- Dynamic UI
- User Authentication
- Store & Write User Data

## Production Deployment

- AWS
- EC2 Server
- VPC Connection
- URL Routing

# Shiny Apps for Business (DS4B 202A-R)



Web Application Development  
**6 Weeks**





# 15% OFF PROMO Code: learninglabs

## R-TRACK BUNDLE

**4-Course Bundle - Machine Learning + Expert Web Applications (R-Track)**

Go from Beginner to Expert Data Scientist & Shiny Developer in Under 6-Months

4 Course Bundle ~~\$1,500~~

**DS4B 101-R: Business Analysis With R**

Your Data Science Journey Starts Now! Learn the fundamentals of data science for business with the tidyverse.

Matt Dancho

**DS4B 102-R: Shiny Web Applications For Business (Level 1)**

Build a predictive web application using Shiny, Flexdashboard, and XGBoost.

Matt Dancho

**DS4B 201-R: Data Science For Business With R**

Solve a real-world churn problem with H2O AutoML (automated machine learning) & LIME black-box model explanations using R.

Matt Dancho

**DS4B 202A-R: Expert Shiny Developer with AWS**

Learn how to build Scalable Data Science Applications using R, Shiny, and AWS Cloud Technology.

Matt Dancho

<input type="radio"/>	<b>Paid Course</b> 15% COUPON DISCOUNT	\$1,596 \$2,356.60
<input checked="" type="radio"/>	<b>12 Low Monthly Payments</b> 15% COUPON DISCOUNT	12 payments of \$149/m 12 payments of \$126.65/m

# Begin Learning Today

[university.business-science.io](http://university.business-science.io)

