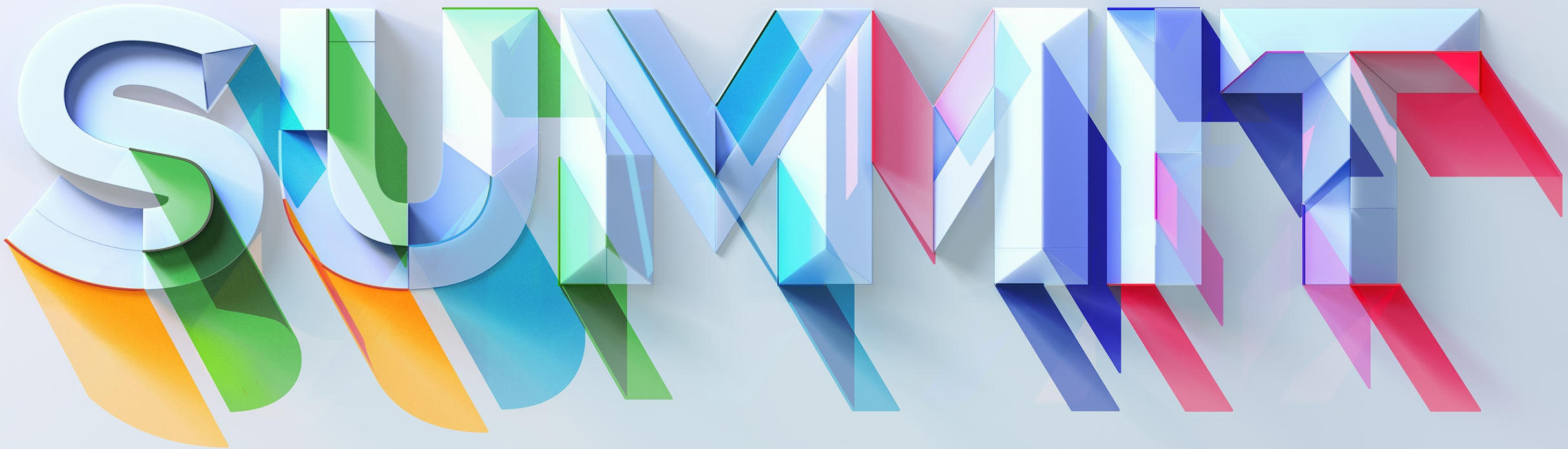
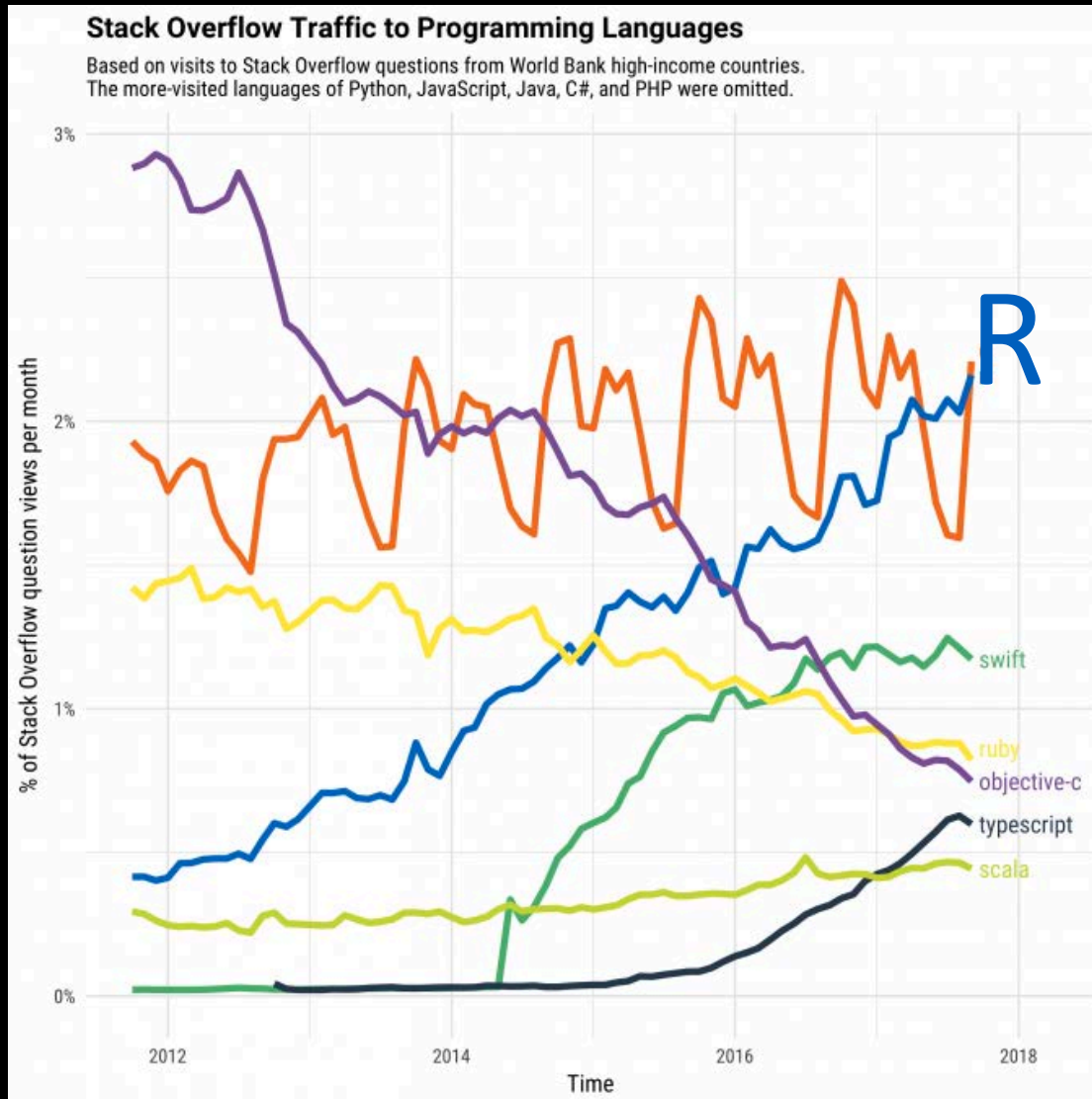




# Advanced Analysis with Adobe Analytics Data Feeds and R

Trevor Paulsen | Principal Product Manager, Adobe Analytics





Source: <https://stackoverflow.blog/2017/10/10/impressive-growth-r/>

- R has been growing in popularity since 2005 (7<sup>th</sup> most popular language worldwide)
- R developers make \$131,205 on average
- In my experience, used by all the best analysts!



+



=



# Trevor Paulsen

Principal Product Manager,  
Adobe Analytics  
tpaulsen@adobe.com



# What is R?





An **Open Source** (Free!\*)

# **Programming Language**

Of, For, and By the **Analyst/Data Scientist**

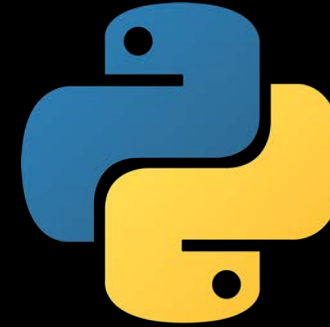
That Readily Connects to **Many Data Sources**

And Has **Robust Sharing/Distribution Options**

\* any open source software is subject to its own licensing



VS

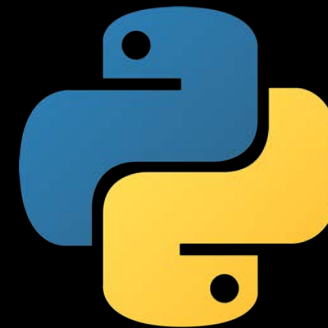


- Comes from academia – typically has the latest and greatest algos
- Best for analysis and insight sharing
- Tons of libraries for just about anything
- Higher learning curve – typically have to know statistics to be proficient

- Comes from the developer community – is very popular
- Best for deploying machine learning to production applications
- Lots of cool packages for analysis
- Lower learning curve – easier to use without statistical background



VS



# Both are awesome!

(but I have a soft spot for R)



# 1

Warm up with dplyr and RStudio

# 2

Connect to the cloud! Azure, AWS, BigQuery, or any SQL DB

# 3

Do awesome stuff, and share it with your whole org

**YOU GET ALL THE CODE!**

# What you'll be able to do after this session

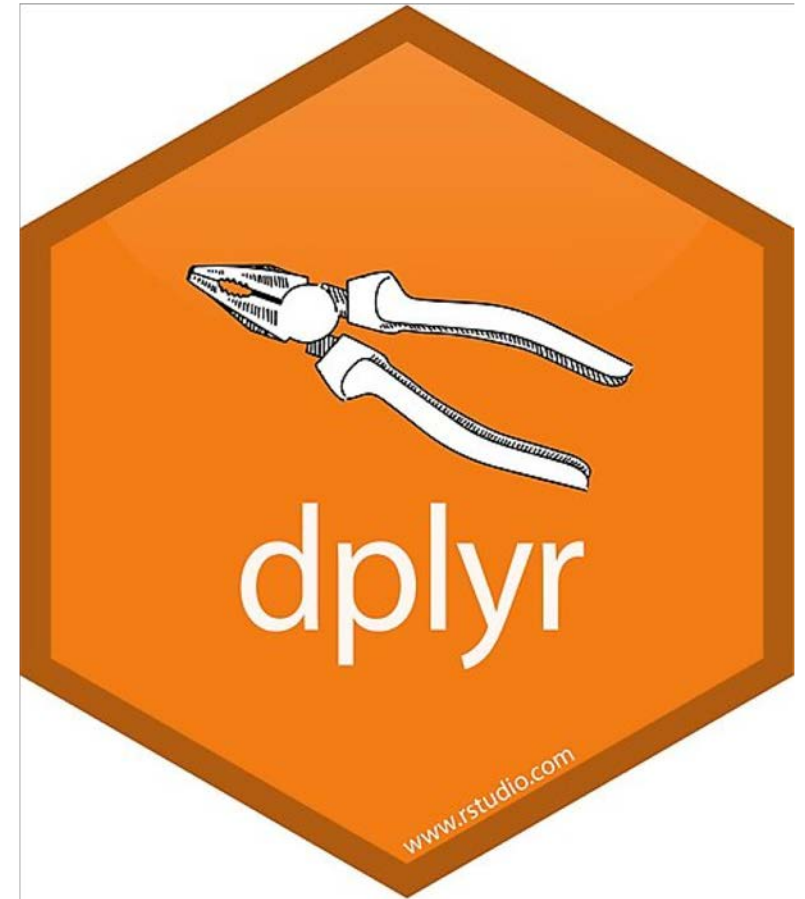
1. Connect R to your cloud or on-premise database
2. Build a cross-device dataset from Adobe Analytics data feeds
3. Cross-device, algorithmic attribution
4. Richly visualize your customer journey using Plotly
5. Create a cluster model, visualize it and share it with Analysis Workspace

# Rstudio, R Notebooks, dplyr, and Data Feeds



# dplyr: R Library for Easy Data Manipulation That Scales

1. Easy to learn syntax
2. Function chaining makes code much smaller/simpler than other languages like SQL or Python
3. Direct connections to external databases!



# What Are Adobe Analytics Data Feeds?

- Raw, hit level data – it's the data Adobe uses to create your reports
- The most granular level of data you can get – they're also really BIG and usually won't fit on your laptop
- Can be difficult to work with – require a lot of transformation







# Workspace



My Favorite Projects

You have not marked any projects as favorites.

Frequently Viewed Projects

- Datafeed Toolbox
- Saturday Night Live
- SNL Starter Demo
- Uber Pickups
- Adobe Analytics + Microsoft Dynamics CRM
- Compression Ratio Test
- Cross Device Measurement
- Sparklyr Attribution Test

Name



Adobe Analytics + Microsoft Dynamics CRM

Saturday Night Live

SNL Starter Demo

Admin Tools Home

Report Suites

User Management

Classification Importer

Classification Rule Builder

Data Connectors

Data Feeds

Data Sources

Dynamic Tag Management

Tag Management

Code Manager

Company Settings

Create New Project

Manage Projects | View Tutorials

Created By

Tags

Trevor Paulsen

Trevor Paulsen

Trevor Paulsen







## Data Feeds

Feeds

Jobs



> Report Suites (0)

28

> Owner (0)

72

> Status (0)

5



Add

Search By Feed Name



FEED NAME

REPORT SUITE



prod\_jun2015\_s3 adbadoben...

Global Production



trial

Global Production



adobe prod

Global Production

# New Data Feed


Cancel

Save

▼ Feed Information ?

NAME

My Data Feed



REPORT SUITE

Datafeed Toolbox ▼

Select Report Suite ▼

X

EMAIL WHEN COMPLETE

tpaulsen@adobe.com

FEED INTERVAL

Daily ▼

DELAY PROCESSING

No Delay ▼

START & END DATES

Feb 27 2018 - Feb 28 2018

☐ CONTINUOUS FEED

▼ Destination ?

TYPE

FTP ▼

HOST

Enter FTP Host (ftp.domain.com/subdomain)


PATH

Enter Path

☐ APPEND REPORT SUITE ID TO PATH


USERNAME

Enter Username




PASSWORD

Enter Password



CONFIRM PASSWORD

Re-Enter Password



☒ REMOVE ESCAPED CHARACTERS

COMPRESSION FORMAT

Gzip

PACKAGING TYPE

Multiple Files

MANIFEST

Manifest File

Select Column Template

Download CSV

AVAILABLE

Add All

Search

accept\_language

aemassetid

aemassetsource

aemclickedassetid

browser

browser\_height

browser\_width

c\_color

carrier

channel

Add

→

INCLUDED COLUMNS

Remove All

Search

1. campaign

2. post\_visid\_high

3. post\_visid\_low

4. hit\_time\_gmt

5. event\_list

# How to Download Your Classifications





# Classification Importer

[Download Template](#)

[Import File](#)

[Browser Export](#)

[FTP Export](#)

## Overview

When you download a Classification file, the file format must follow the Classifications guidelines. Here you can download a template to use as a guide for data import. Click on the "Download" button below to immediately download the template and comments for the selected base attribute and report suite.

## Download Template

Select Report Suite:	<input type="text" value="Datafeed Toolbox"/>
Data Set to be classified:	<input type="text" value="Campaign"/>
Export Numeric 2 <i>beta</i> :	<input checked="" type="radio"/> No <input type="radio"/> Yes
Encoding:	<input type="text" value=" &lt;Default&gt;"/>

[Download](#)

[Admin Tools Home](#)

[Report Suites](#)

[User Management](#)

[Classification Importer](#)

[Classification Rule Builder](#)

[Data Connectors](#)

[Data Feeds](#)

[Data Sources](#)

[Dynamic Tag Management](#)

[Tag Management](#)

[Code Manager](#)

[Company Settings](#)



# Classification Importer

[Download Template](#)[Import File](#)[Browser Export](#)[FTP Export](#)

## Source

Select Report Suite:

Available Report suites

Report Suite 1  
Report Suite 2  
Report Suite 3  
Report Suite 4  
Report Suite 5

&gt;&gt;

&lt;&lt;

Selected Report Suites

Datafeed Toolbox

Data Set to be classified:

Campaigns

## Filter Data

Select Number of Rows:

☒ All

☐ Limit Data Rows to

Filter by Date Received:

Download rows that were received between Jan 2015 and Oct 2017

Note: This applies to rows received via Classifications or through standard collection methods.

Apply Data Filter:

Download rows that match all of the following selected filters

☐ Rows where: Campaigns = (Use \* as a wildcard)

☐ Rows with empty column: --All Columns--

Date Filter:

☒ No Filter

☐ Only Active Campaigns (Includes all open campaigns and those with no end date assigned)

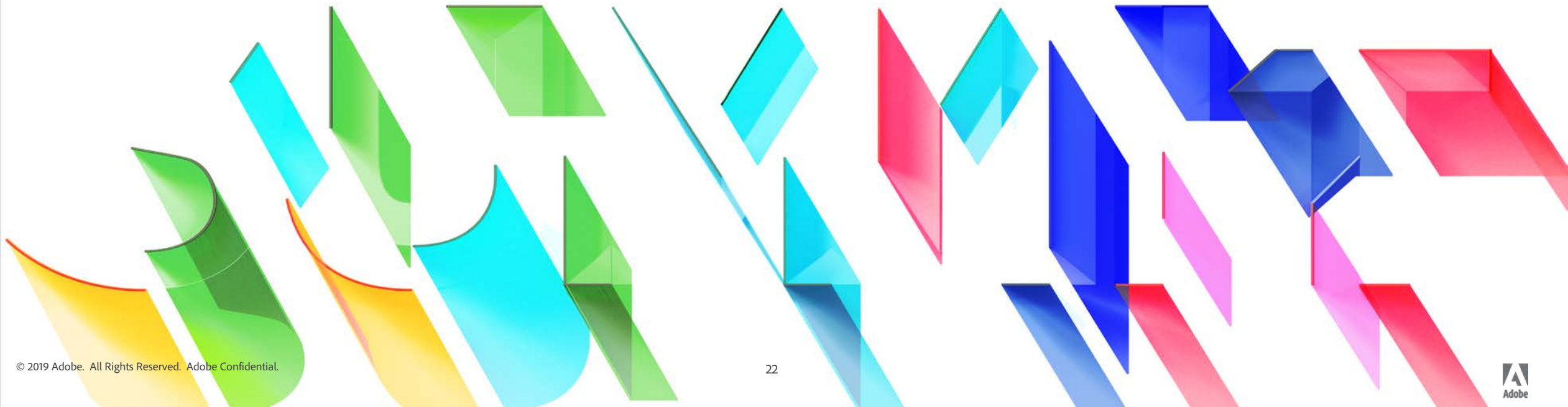
☐ Show ONLY Campaigns that:

☐ Began: Between 7 October, 2017 7 October, 2017

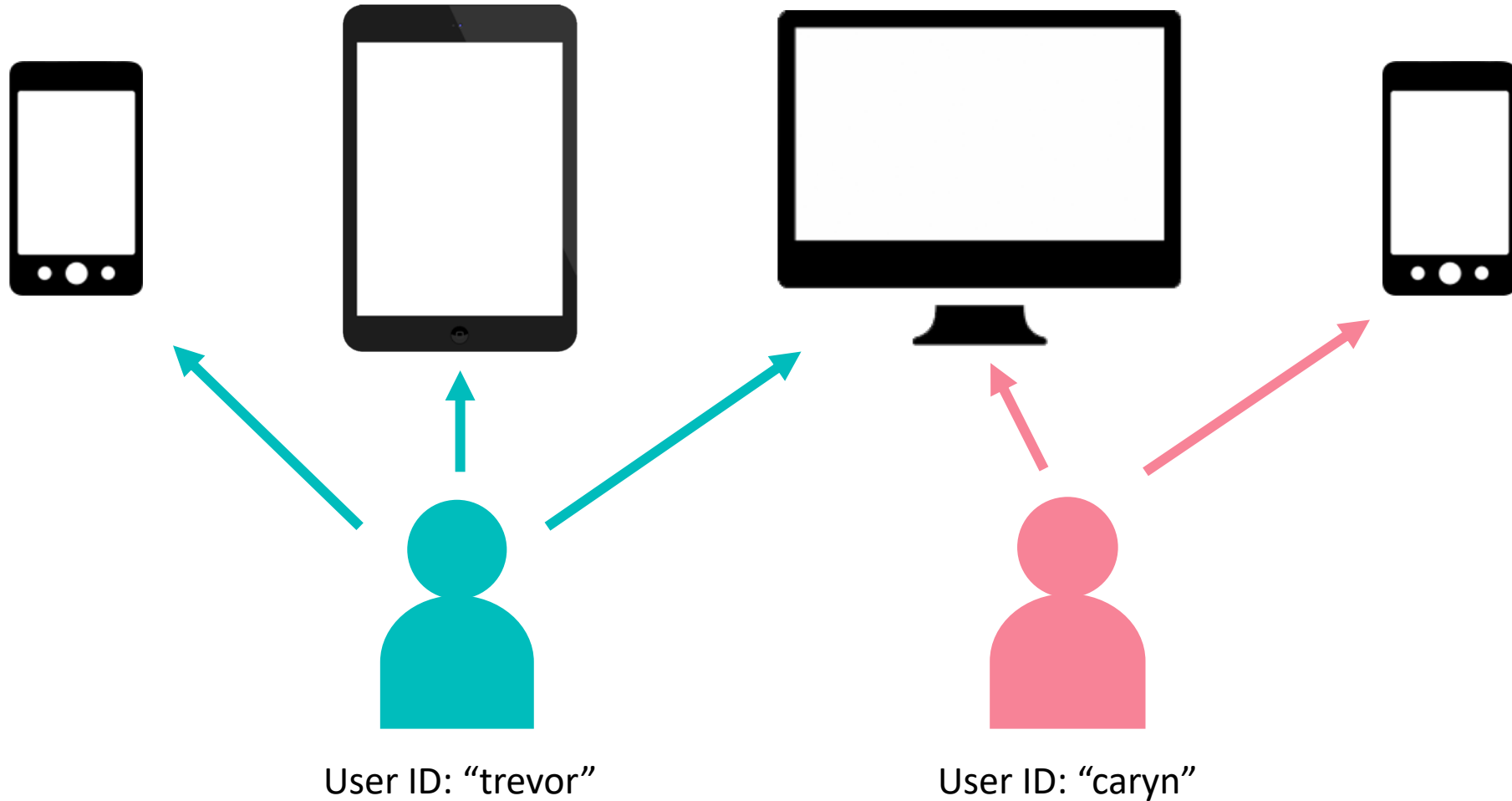


# DEMO

# Going Cross-Device



# Cross-Device Is A Tricky Beast!



# DEMO

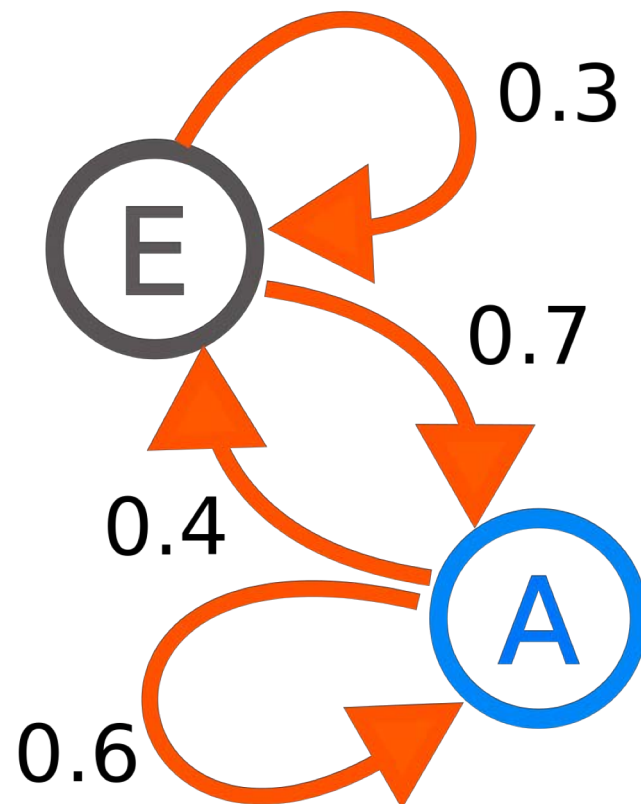
# Algorithmic (Markov) Attribution



# Markov Chains: A Cool Way to Do Attribution



Andrey Markov  
source: Wikipedia



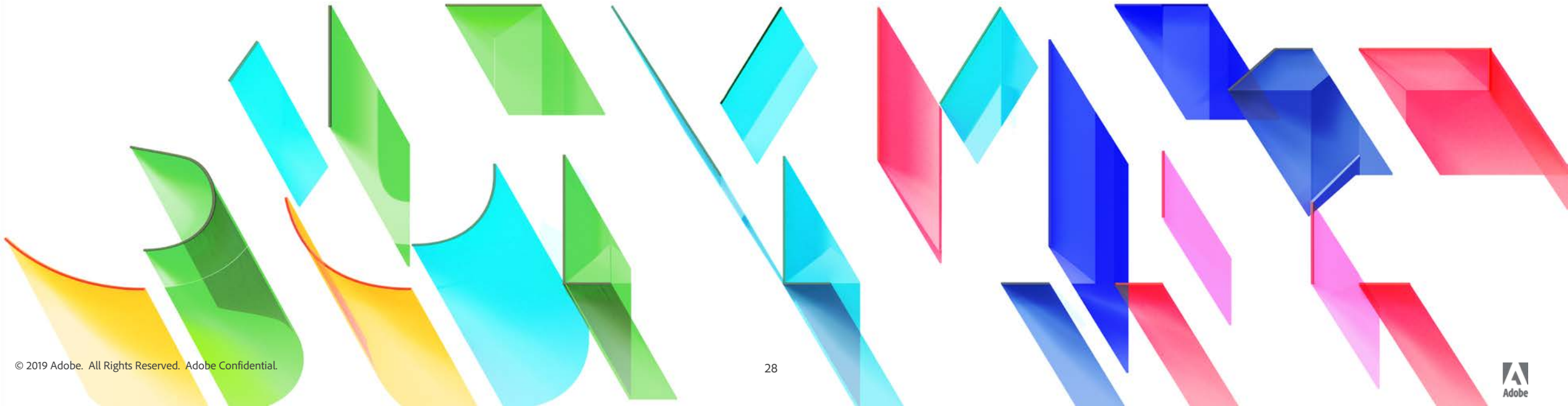
Basic idea:

- Figure out how often people move from one channel to another before converting.
- See what would happen if you removed a channel
- If it had a big impact on conversion, it should get a lot of attribution credit!



# DEMO

# Visualizing the Customer Journey



# Plotly: An Open Source, Cross Platform Viz Library



- Plotly creates interactive plots that can be easily shared
- Works great with ggplot
- Works with R, Python, node.js, Scala, etc.

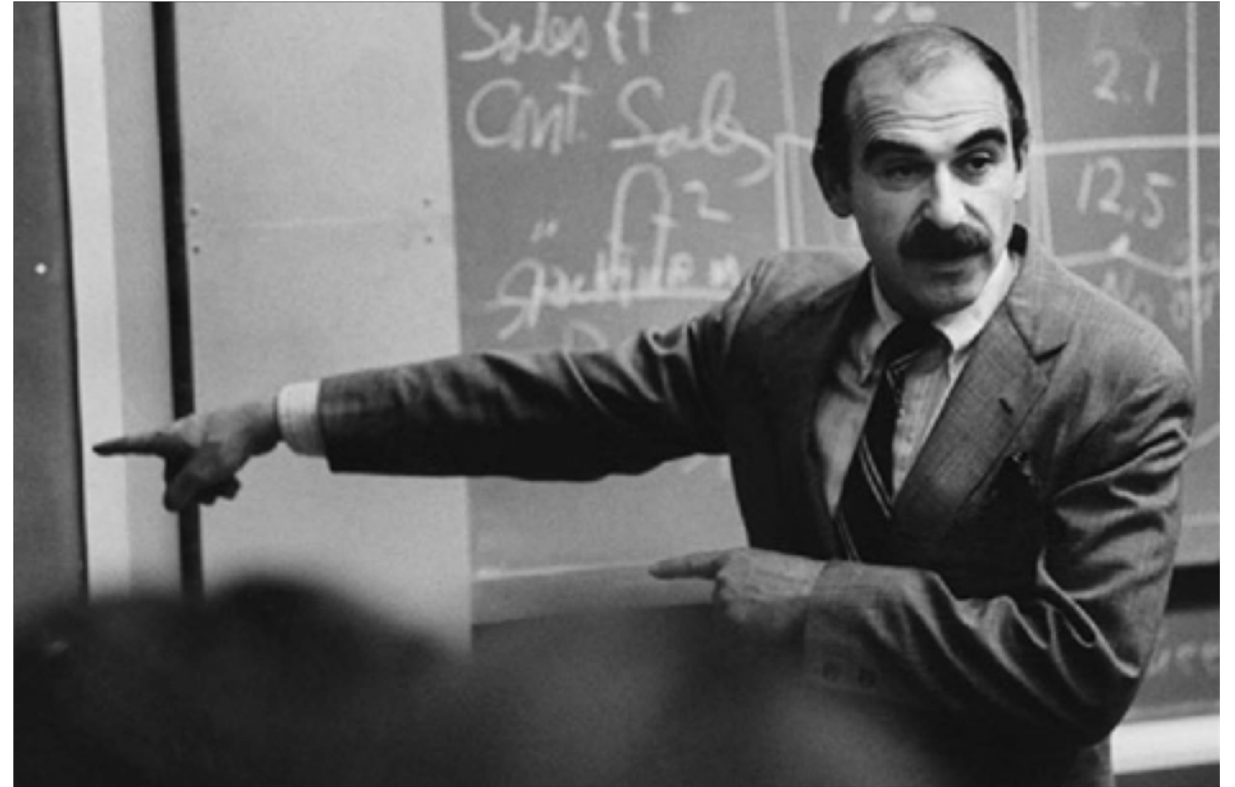
# DEMO

# Clustering Your Users Or Customers



# Clustering Your Users/Customers

“If you’re not thinking segments, you’re not thinking.”



Theodore Levitt  
Harvard Economist



# DEMO

# Make Your Clusters Actionable in Workspace



1

3



Analytics



Social



Audience Manager



Campaign



Experience Manager



Media Optimizer



Primetime



Target



Activation



Assets



Exchange



Feed



Mobile Services



Profiles & Audiences

2

### Name & Description

Clustering Algorithm Classifications

Using a clustering of visitors in R, and bringing them into Adobe Analytics for reporting

### Alias ID

To track data, enter a unique ID that will be used in your Customer Attribute Source Code. ID should be unique, all lowercase with no spaces. Click [here](#) for instructions.

aaid



### File Upload

Your data file must comply with the file [Upload Requirements](#) and must not exceed 100MB. If your file is too big or you have data that will need to be uploaded on a recurring basis you can [FTP](#) your files instead.

Drag and drop CSV/ZIP/GZIP file here, or [Browse](#)

File Name	Date Uploaded	Upload Type	Records	New Records	Updated Records	Status	
All Data	-	-	629998	314999	0	-	✓
cluster_mapping.csv	Sun Apr 02 05:27:49 UTC 2017	http	314999	0	0	COMPLETE	
cluster_mapping.csv	Sun Apr 02 05:26:15 UTC 2017	http	314999	314999	0	COMPLETE	

### Configure Subscription

+ New

Solution	Attributes Mapped	
× Analytics	1	✓
× My Report Suite	1	

### Activate

☒ Active ☐ Inactive



# DEMO

Summit 2019 Clustering

☆

Project

Edit

Insert

Components

Share

Help

Adobe Summit 2019 (2)

▼

Search Components

DIMENSIONS

Day

Month

user\_id

VI\_V2

channel

Show All

>

METRICS

Unique Visitors

Occurrences

orders

revenue

VI

Show All

>

SEGMENTS

Purchasers

Cluster Comparisons

▼

×

cluster

No filter

▼

Drop a Segment Here (or any other component)

Jan 5 2018 - Mar 7 2018

Freeform Table

	Unique Visitors	Revenue/Visitor	revenue	internal_searches	Internal Searches/Visitor
cluster					
Page: 1 / 1 Rows: 50 1-12 of 12	25,620	\$423.69	10,854,826.00	298,400.00	↓ 12
1. 0	8 0.0%	\$0.00	0.00 0.0%	4,600.00 1.5%	575
2. 3	121 0.5%	\$1,201.40	145,370.00 1.3%	32,000.00 10.7%	264
3. 10	61 0.2%	\$6,794.43	414,460.00 3.8%	10,100.00 3.4%	166
4. 4	237 0.9%	\$0.00	0.00 0.0%	31,200.00 10.5%	132
5. 11	155 0.6%	\$43,115.34	6,682,877.00 61.6%	19,500.00 6.5%	126
6. 2	869 3.4%	\$2.47	2,148.00 0.0%	59,800.00 20.0%	69
7. 9	202 0.8%	\$14,361.04	2,900,931.00 26.7%	11,200.00 3.8%	55
8. 7	33 0.1%	\$0.00	0.00 0.0%	1,400.00 0.5%	42
9. 8	124 0.5%	\$5,718.06	709,040.00 6.5%	3,600.00 1.2%	29
10. 6	102 0.4%	\$0.00	0.00 0.0%	2,300.00 0.8%	23
11. 5	158 0.6%	\$0.00	0.00 0.0%	1,700.00 0.6%	11



# Takeaways

# 1

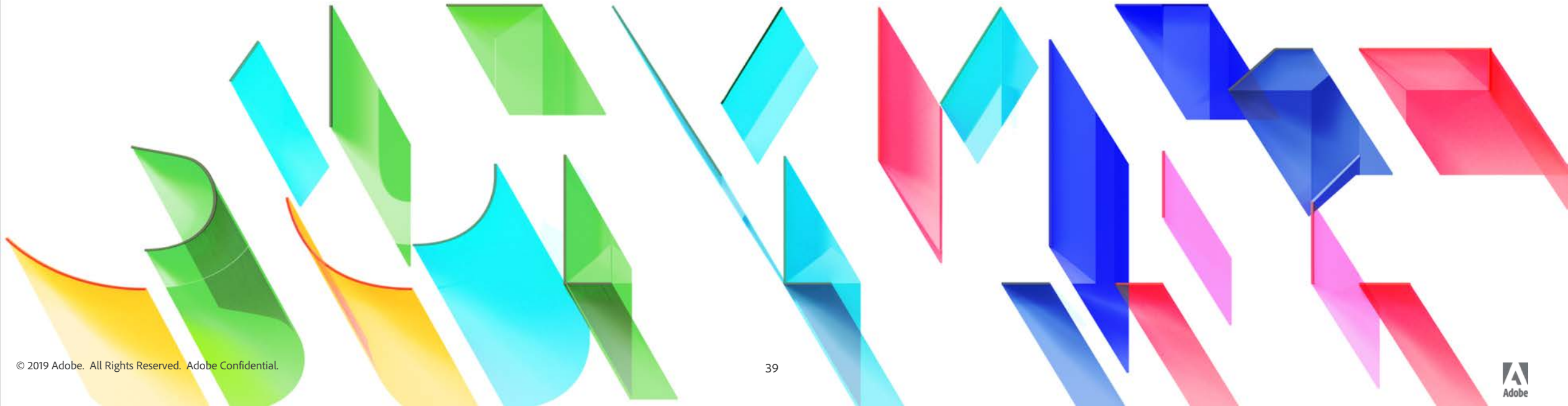
Get your data feeds into your data lake and start using R!

# 2

Connect to the cloud! Azure, AWS, BigQry, Spark

# 3

Do awesome stuff, and share it with your whole org



# Resources

1. Code samples from this session: <https://github.com/admoseremic/summit2019>
2. My blog: <http://datafeedtoolbox.com>
3. Getting started with R: <http://r-marketing.r-forge.r-project.org/>
4. Getting started with RStudio: <http://rstudio.com>
5. Getting started with Plotly: <https://plot.ly/>

# Take the Survey for a chance to win!

(Survey section of the mobile app)



**\$10** Starbucks Card



Day 1  
Bash  
Experience



Day 2  
Signed Sports  
Memorabilia



Day 3  
Bose  
Home Speaker

# Experience Success expert zone at the Adobe booth.

- Talk with Adobe's top experts in digital experiences
- Learn about Adobe's Digital Performance Services
- Meet with customer support teams
- Check-in for your personalized assessment

VISIT US AND  
*Enter*  
TO *WIN*  
**AN iPad PRO**  
& OTHER PRIZES

[Adobe.com/go/experience-success](https://adobe.com/go/experience-success)





If you fill out the survey, I'll send you bonus code including:

- Shapley Value Attribution
- Building a propensity model
- Search term text mining visualization

# Q & A

Trevor Paulsen

tpaulsen@adobe.com

Twitter: @trevorwithdata

LinkedIn: /trevorpaulsen/







ADOBE SUMMIT | MARCH 26–28 LAS VEGAS



**Adobe**