Unit 4 (Week 5): Content Marketing & Explanatory Analytics (Part 1)

Course: Business and Marketing Analytics

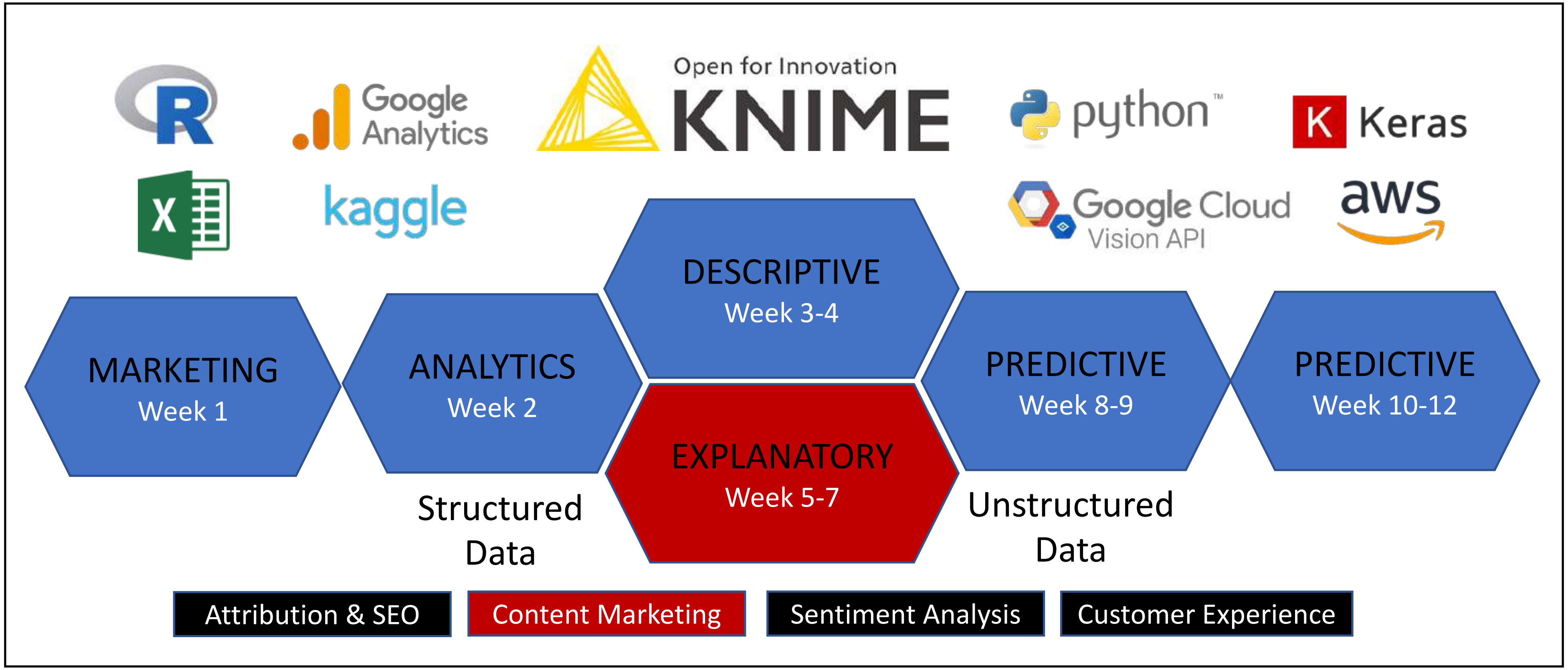
Francisco Villarroel Ordenes

October, 2023

LUISS

Recap Unit 3

- **Descriptive Analytics**
- Channels and Attribution Analysis
- Google Analytics Tool and Conversion
- Onsite and Offsite SEO (Keyword Analysis)

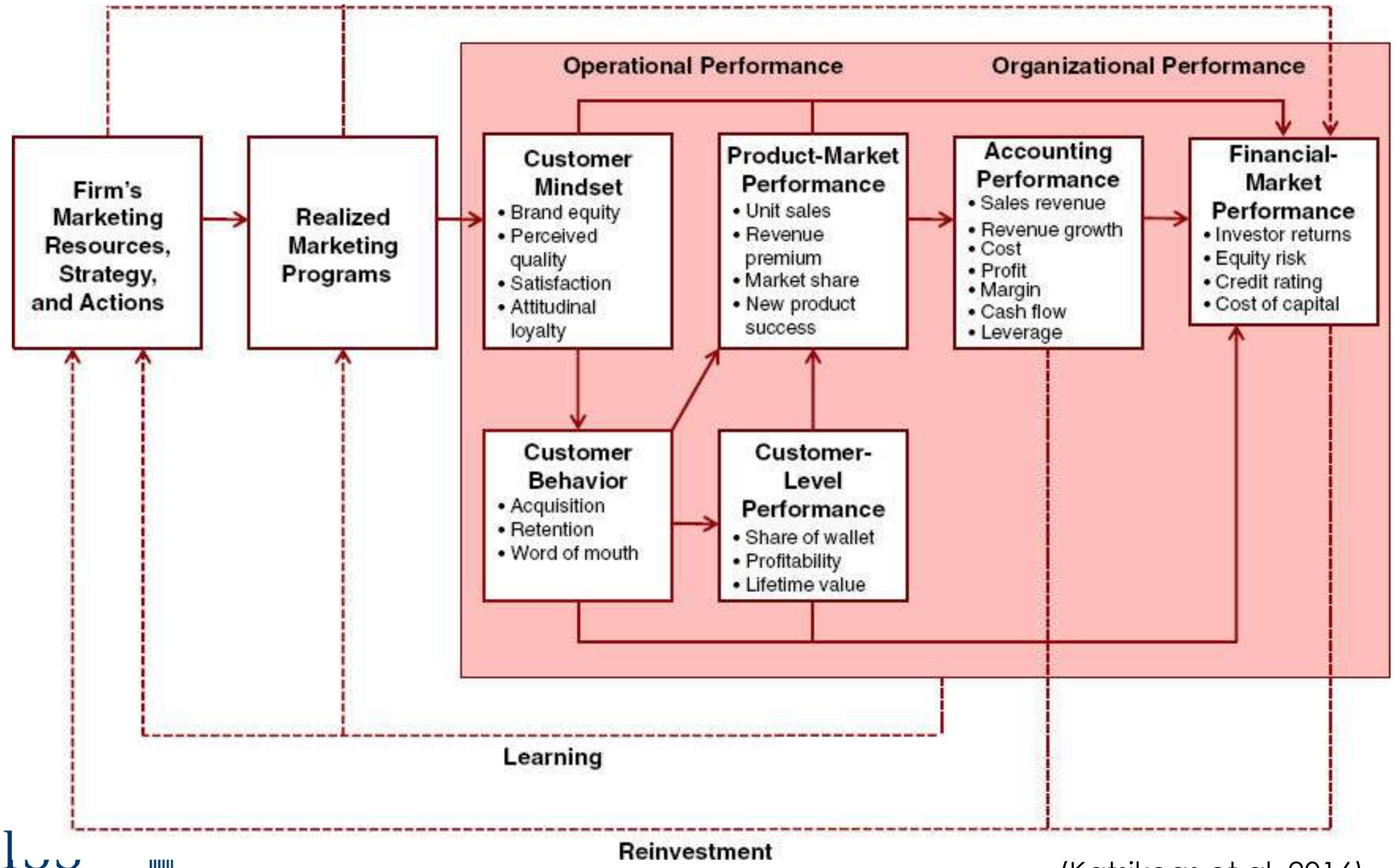


BUSINESS & MARKETING ANALYTICS FRAMEWORK

Unit 4 Objectives

- **Part 1 (Today)** Introduce the context – Integrated Marketing Communications, Content Marketing, e-WOM, Social Media Marketing
- **Part 2 (Tomorrow)** Focus on explanatory analytics
- Different types of RQ and/or hypothesis that can be tested through explanatory analytics
  - Main effect (What)
  - Moderation (When)
  - Mediation (How)
- Implement explanatory analytics models using field social media data

Context for Explanatory Analytics (A Performance Perspective)



(Katsikeas et al. 2016)

Unit 4

We will focus on the impact of marketing actions on social media word of mouth (engagement)

INTEGRATED MARKETING COMMUNICATION (Batra and Keller 2016)

Overview of relevant topics in Marketing Communication & implications

DIGITAL CONTENT MARKETING (Hollebeek and Macky 2019)

Framework for new types of content beyond Advertising

SOCIAL MEDIA CONTENT (Villarroel O. et al. 2019)

Analytics in Social Media Brand posts

What is the purpose of Integrated Marketing Communications? Why firms do it?

Move Consumers through Funnel | Deep and positive interaction

Short Term

Awareness, **I**nterest, **D**esire, **A**ction (**AIDA**)

Engagement (Cognitive, Emotional, Behavioral)

Long Term

Build Brand Equity
"differential preference and response to marketing effort that a product obtains because of its brand identification (Datta et al. 2017)

Consumer-Based Brand Equity

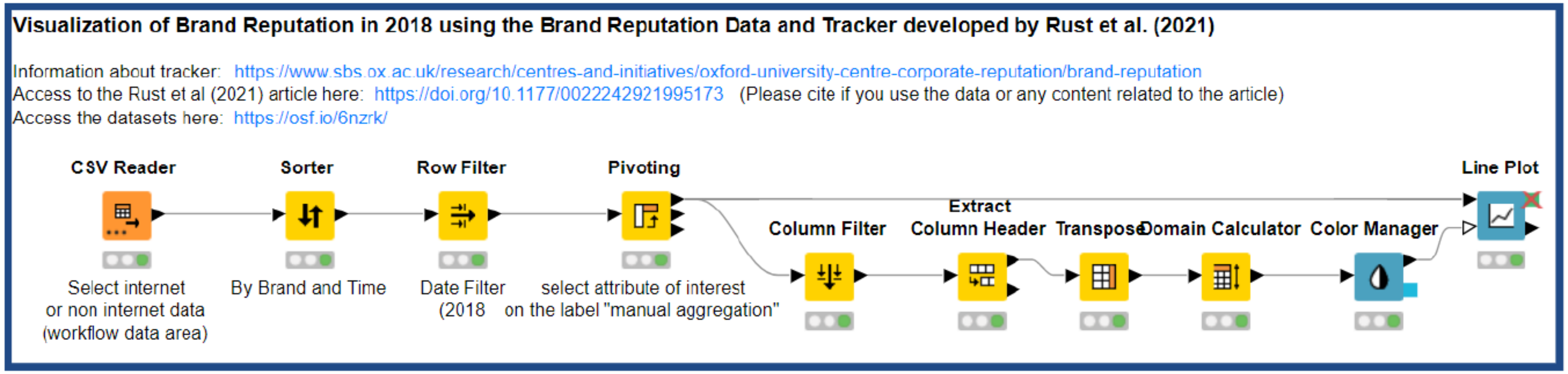
Sales-Based Brand Equity

Relation between Customer Based Brand (CBBE) Equity and Sales Based Brand Equity (SBBE)

Datta et al. 2018

New models to measure Brand equity

- **Brand Reputation Tracker** https://kni.me/w/-ClH0lgtN3fw14G-



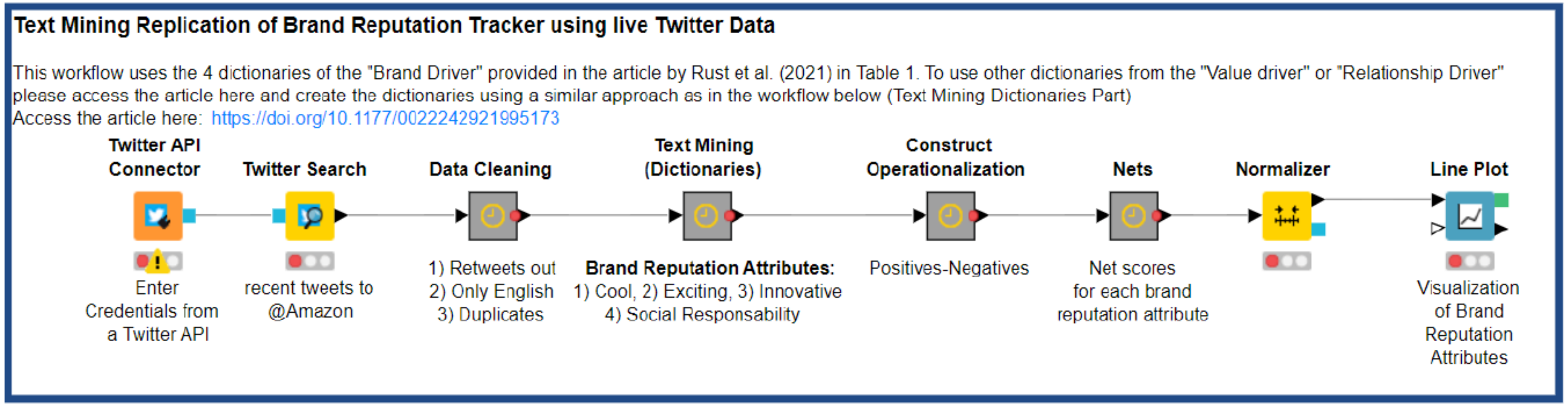


Table W1. A Comparison with Existing Brand Measures.

	Granular level	Aggregate level	Conceptual source	Customer's voice/ Stakeholders	Availability
Brand Reputation Tracker	Weekly, monthly, quarterly	• 100 global brands • At the overall brand reputation, driver, and sub-driver levels • Brand sentiments (positive and negative)	Rust et al.'s (2000, 2004) customer equity framework	• Stakeholders' own voice on Twitter • Stakeholders, (current and prospective) customers, employees, and investors	Publicly available
Y&R's BrandAsset Valuator (BAV)	Quarterly, annually	• Brand strength (relevance and differentiation) and brand stature (esteem and knowledge)	Keller's (1993) customer-based brand equity (CBBE) model	• Consumer perception surveys • Customers	For paid clients only
YouGov's BrandIndex	Daily, annually	• An overall ranking of 25 global healthiest brands	Calculate overall brand health index (net of positive and negative feedback) based on 6 metrics: impression, quality, value, satisfaction, recommend, and reputation	• Consumer's brand perceptions based on consumer panel interview • North America only • Consumers	Brand health index released annually, and daily data are only available for paid clients
BrandZ's top 100 most valuable global brands	Annually	• An overall ranking of 100 brands	Calculate the proportion of a brand's contribution to its firm's financial value	• Firm financial value based on expert judgment and calculation • Brand contribution based on consumer interviews • Customers	Publicly available
Interbrand's Best Global Brands	Annually	An overall financial brand value ranking for 100 brands	• A firm's overall financial returns • The brand's contribution to the financial returns • Brand loyalty	• Expert evaluation based on financial data, consumer goods data, and text analytics and social listening • Stakeholders	Publicly available
Forbes' World Best Brands	Annually	An overall financial value ranking for 100 brands	A brand's contribution to the firm's financial value	• Financial revenue and earnings based • Investors	Publicly available

Integrated Marketing Communication Strategy (Batra and Keller 2016)

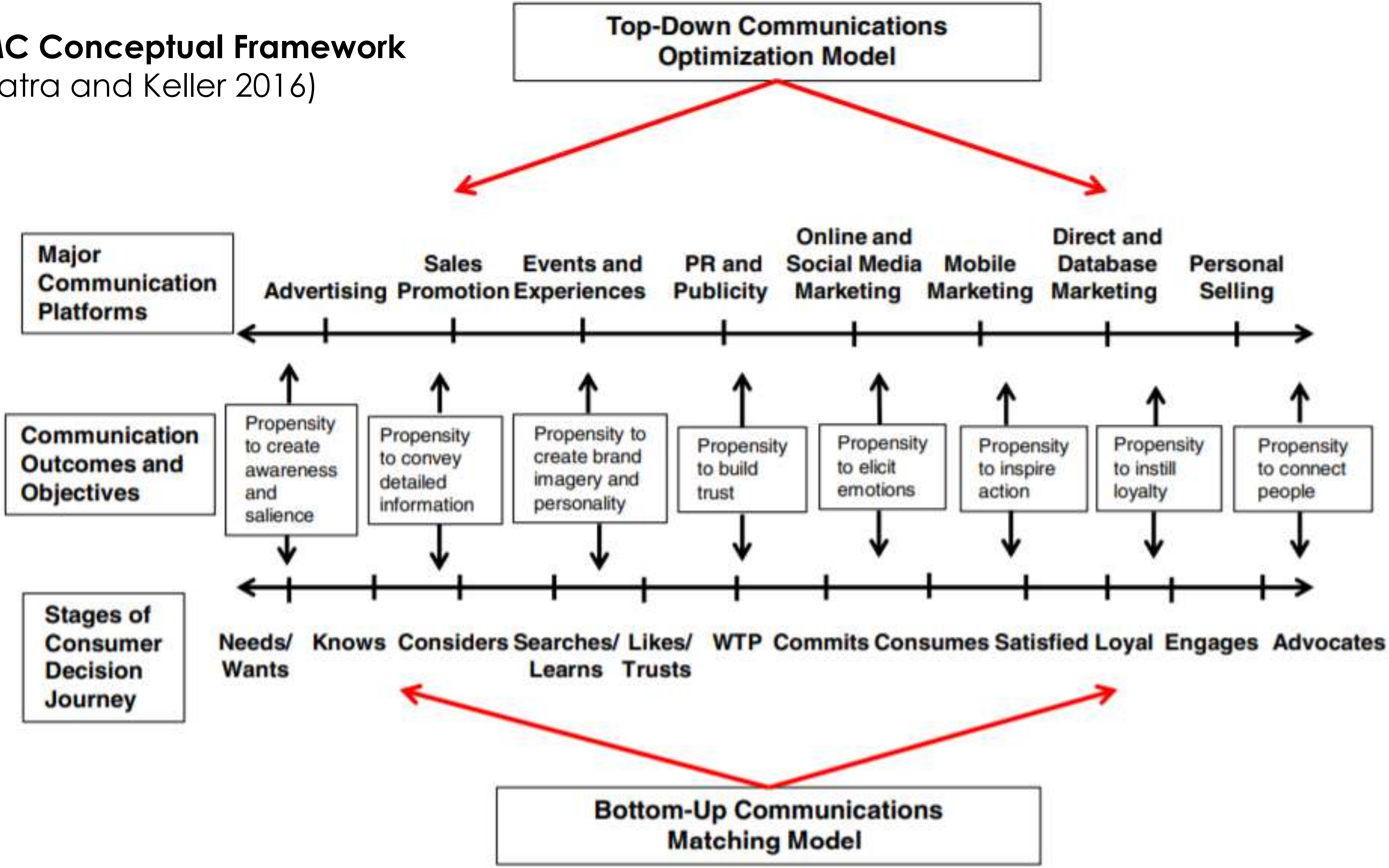
- **Consistency:** To facilitate learning and induce action, the exact same persuasive message (e.g., content about brand attributes) can benefit from being reinforced in different ways across different communications
  - Volvo's key "Safety superiority message" (across all channels)
- **Complementarity:** different communication options have varied strengths and weaknesses, which can meet different brand-related information needs (e.g., awareness, action) for consumers and, thus, complement each other.
  - Volvo's Sponsoring (e.g., ocean race) -> brand visibility & salience
  - Volvo's Sales promotion -> push safety concerned consumers to take action
- **Cross-effects:** Communication effects from consumer exposure to one communication option can be enhanced when consumers have had prior exposure to a different communication option (e.g., spillovers: paid search and search)
  - Good feelings & awareness triggered by ocean race sponsorship may later increase predisposition to consider the brand

Integrated Marketing Communication (Batra and Keller 2016)

Communication outcomes depend on the Customer Journey Stage:

- *Create Awareness & Salience:* Foundation of a brand
- *Convey Detailed Information:* Why is better than competition?
- *Create Imagery and Personality:* Intangible benefits, human-like traits
- *Build Trust:* Credibility, Expertise, Likability, Authenticity facilitates message acceptance
- *Elicit Emotions:* Brands can increase their perceived value by hedonic associations
- *Inspire Action:* when the objective is making the conversion in consumers already favorable predisposed about the brand
- *Instill Loyalty:* After consumption, shape satisfaction to avoid customer deflection
- *Connect People:* For WOM & advocacy to occur, the consumer needs to **engage & interact frequently** with the brand and develop a sense of "brand love"
    - Send consumers the types of brand messages that would motivate them to pass these along

IMC Conceptual Framework
(Batra and Keller 2016)



Other popular overviews of integrated marketing communication

- "Digital Media" Strategy

Paid Media	Owned Media	Earned Media
Social Media Advertising	Website	Social Media Sharing
Paid Search Marketing	Social Media Profiles	Direct Traffic
Display Advertising	Email Marketing	Search Engine Optimization
Affiliate Marketing		Press Coverage

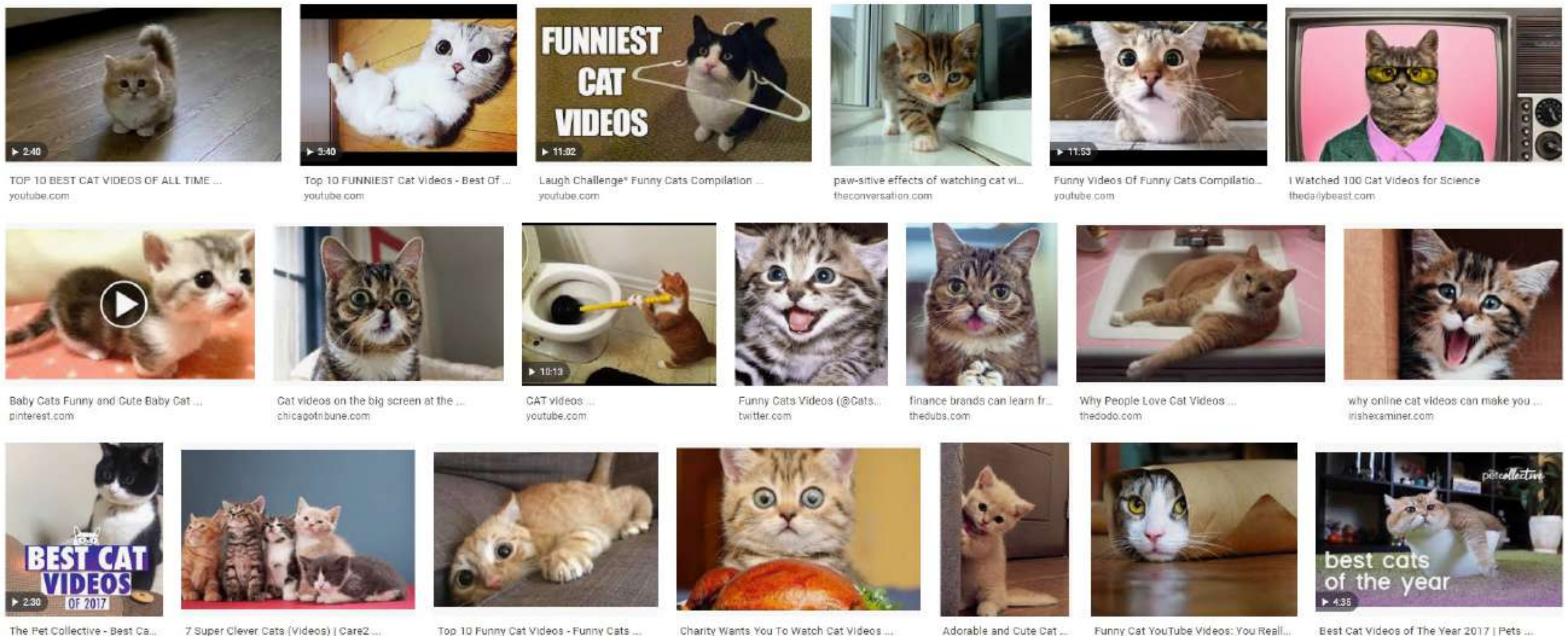
A "balanced" marketing portfolio focuses on all three areas for generating traffic and revenue. What could be the cost of unbalanced?

E-Word of Mouth (eWOM) (Batra and Keller 2016)

- *“If it doesn’t spread it is dead”* (H. Jenkins)
- E-WOM is a large field of research.
- Brands need to make it easy to get their consumers to recommend the brand to others, through Facebook likes, (re)tweets, viral branded content, Instagram comments, and so on
- Through such media tactics, a brand not only can reach many more potential consumers but also can do so in credible, low-key ways that are less likely to evoke hostility and skepticism.

Why do people share things?

[Review the top viral videos of all time here.]

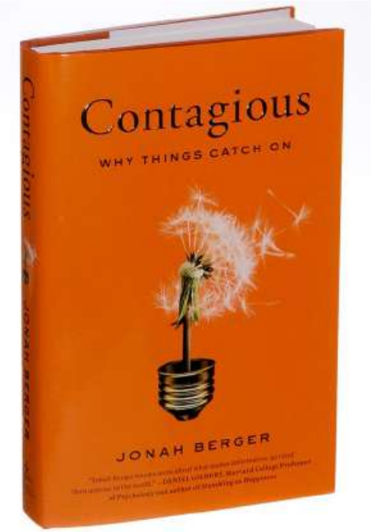


Virality

- Why is it important?
- McKinsey says that WOM generates **twice as many sales as paid advertising** in categories as diverse as skincare and mobile phones
- Why is it so much more effective than advertising?
- We tend to be **friends with people who might like the same products or services that we do**. If you like a product, it is likely that you know more people who would like it too.

How to craft contagious content?

- How can you create content that will get shared, regardless of whom is producing it.
- **S**ocial Currency
- **T**riggered
- **E**motion
- **P**ublic
- **P**ractical Value
- **S**tories

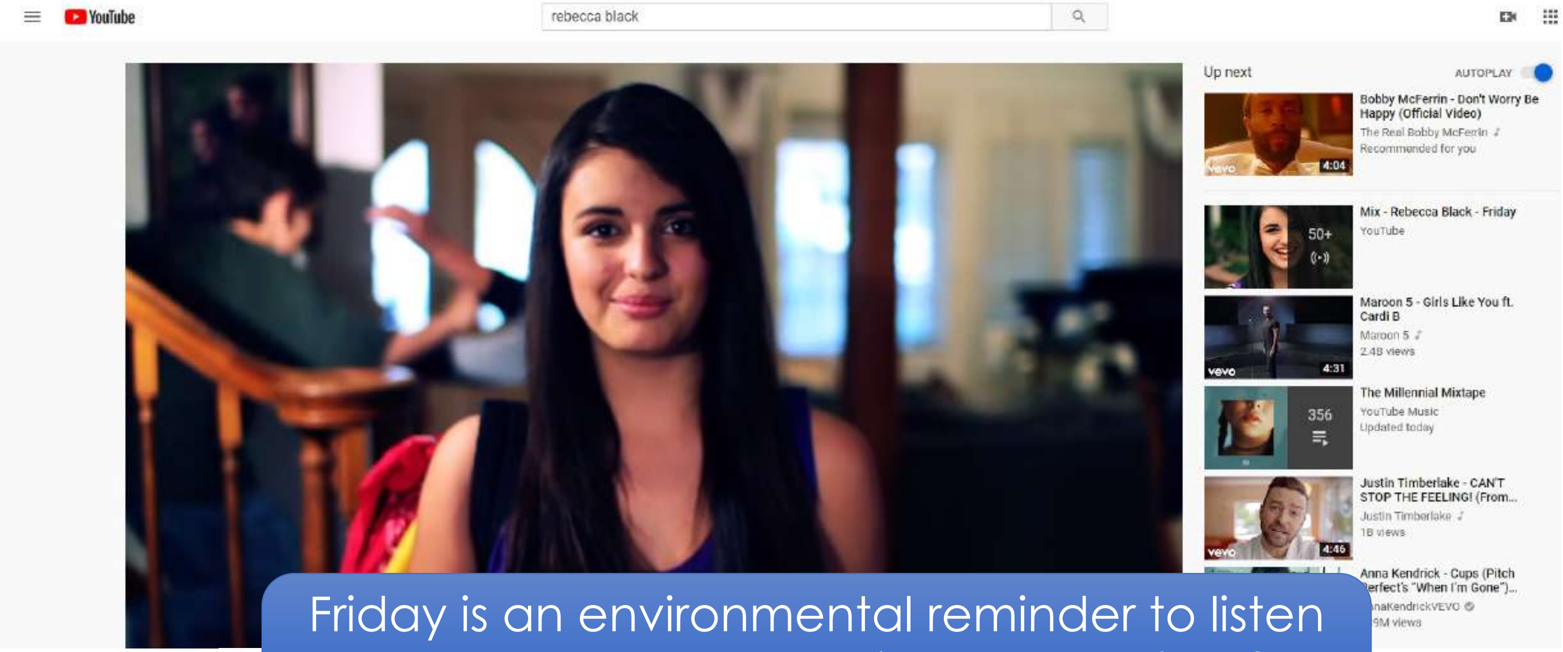



eat me
CRIF ★ DOGS
EAT
3AM
APTS FOR RENT

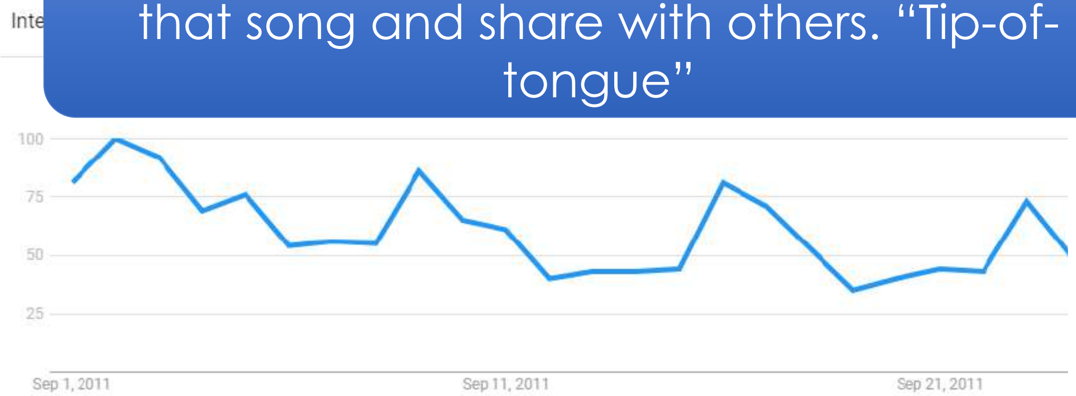
"Please don't tell"

- Secret bar, inside a hotdog restaurant
- Almost impossible to get a reservation
- How did they do it?
- They made themselves "secret" and what you do after someone tells you a secret?
- **"Social Currency"**
- **Knowing this secret makes the information look important, and makes you look good by sharing.**

Triggers



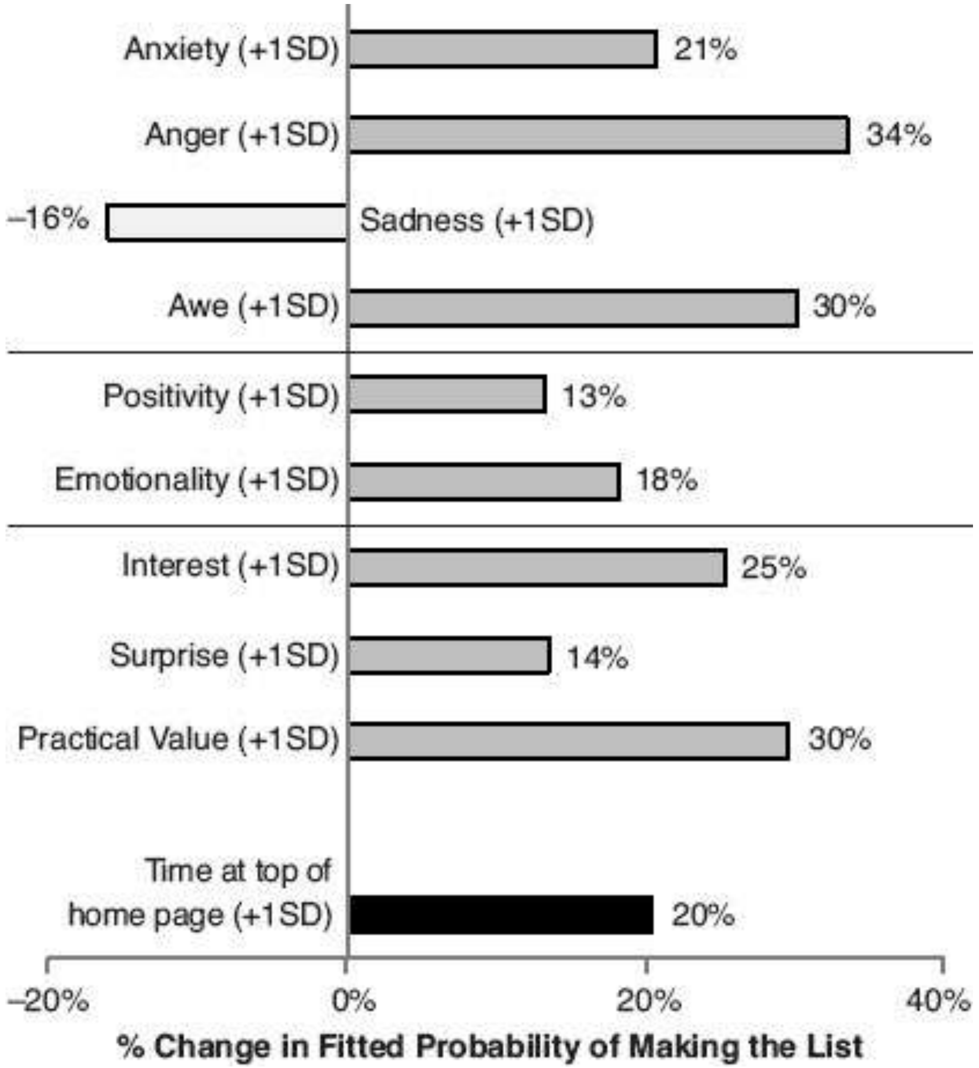
Friday is an environmental reminder to listen that song and share with others. “Tip-of-tongue”




General Mills
FAMILY SI
2 pack
PER 1 CUP SERVING
Cheerios
CAREFULLY
SELECTED OATS
THAT CAN HELP
LOWER
CHOLESTEROL*
23
Toasted Whole Grain Oat Cereal
NET WT 1 LB 5 OZ (21 OZ) (595g)

Emotions

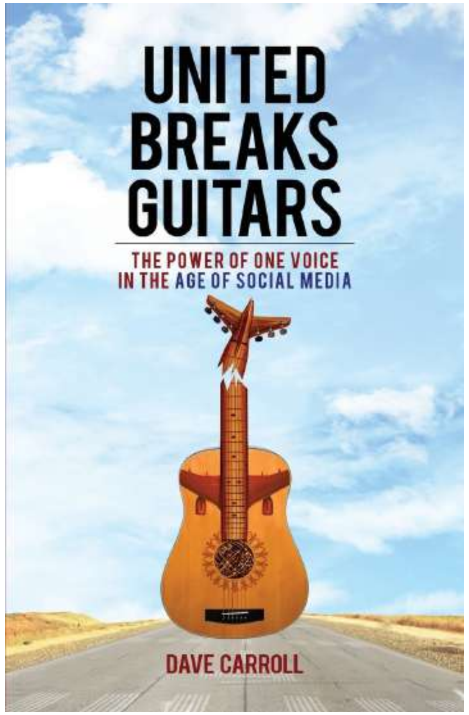
- New York Times articles published over a three-month period
- Does any emotion gets shared?
- Positive or Negative?
- What other element of emotionality is important
- Content that evoke high arousal positive (Awe) or negative (anxiety or anger) is more viral



Story

- Stories carry information and act like a troy horse for the brand that is inside

E-WOM – Negative Consequences



Dave Carrol a musician from Halifax traveled from Halifax to Nebrasca Video costed $150 & It is estimated cost for United was 180 billion!

Part 2: Content Marketing

What is Content Marketing (Hollebeek and Macky 2019)

- **Content Marketing vs. Advertising?**
- **Digital Content Marketing (DCM):** DCM as the creation and dissemination of relevant, valuable brand-related content to current or prospective customers on digital platforms to develop their favorable brand engagement, trust, and relationships (vs. directly persuading consumers to purchase).
- **DCM main Advantage:** It might result in more engaged audiences at a lesser cost!; Consumers are increasingly skeptical about ads.

Content Marketing (Hollebeek and Macky 2019)

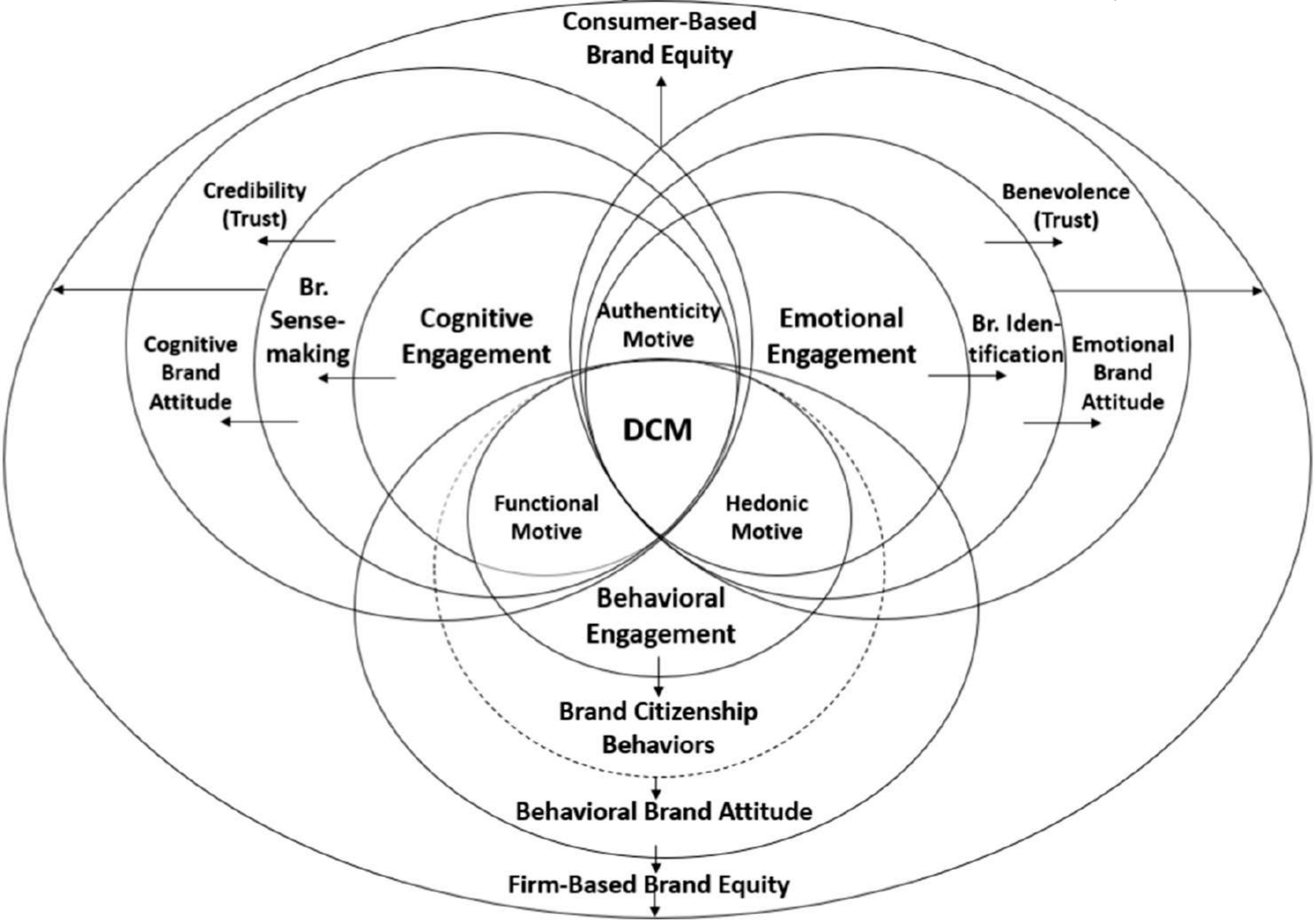
- Content formats disseminated on these platforms include
  - newsletters,
  - ezines,
  - podcasts,
  - live streaming/video,
  - quizzes,
  - whitepapers,
  - infographics,
  - downloadable templates or checklists,
  - case studies,
  - guides,
  - virtual conferences,
  - content hubs,
  - webinars, and
  - longform content (i.e., free content available to subscriber)
  - **social media posts (text, images, video)**

Content Marketing (Hollebeek and Macky 2019)

Four leading premises for DCM:

1. DCM requires a firm's paradigmatic **shift from selling to, to helping consumers** by offering them relevant, valuable content free-of-charge.
2. Designed to build and maintain consumers' long-term **engagement, trust, and relationships** (rather than ads aiming to sell directly). Differently from native advertising, tries to genuinely add value to its audience
3. DCM is based on the premise of **consumer consent, permission**, or opt-in (rather than interrupting)
4. DCM "earns its audience by **offering something of value** [to consumers]"

Figure 1(Hollebeek and Macky 2019): Framework outlines the process by which DCM creates consumer and firm value through a series of consumer-based subprocesses



INSIDE-OUT READING

1.Uses & gratification *perspective*: consumers' functional, hedonic, and authenticity motives drive the decision to interact with DCM. Hybrid motives too

2.Interactions motives *prompt engagement*: Cognitive(functional-authentic), Hedonic (emotional-authentic) and Behavioral (functional-hedonic).

3. *Brand outputs* (*1st* tier): Brand sense-making (cognitive), brand identification (emotional), brand citizenship (behavioral)

4.Brand Attitudes *(2nd tier): Cognitive (Credibility), emotional (benevolence), behavioral.*

5. Final Output: *Consumer and Firm based Brand Equity*.

*While these concepts will sequentially emerge, focal preceding and subsequent concepts with coexist.

Framework

- **Level 1 (Motives)**
  - Functional: use of instruments such as utilitarian scales (practical-impractical)
  - Hedonic: use of hedonic measurement scales (e.g., interesting=boring)
  - Authenticity: Three components (continuity, integrity and symbolism). To symbolism authenticity: [Viewing this content I] feel like [my]self.
- **Level 2 (Consequences – Engagement)**
  - Cognitive: interaction related to brand thought and mental elaboration; brand sense-making
  - Emotional: consumer's degree of positively valenced interaction; brand identification
  - Behavioral: consumer's level of energy, effort, and time spent on a brand; brand citizenship (beyond expected behaviors such as buying; to unexpected behaviors such as positive e-WOM)
- **Level 3 (Second tier consequences)**
  - Brand Trust: 2 dimensions: Credibility(promise will be kept) and Benevolence (act on best interest)
  - Brand Attitude: consumer's "psychological evaluation of a brand with some degree of (dis)favor.
- **Level 4 (Third tier consequences)**
  - Consumer-based brand equity: differential effect of brand knowledge on consumer response to a brand. Mainly affected by cognitive and emotional engagement.
  - Firm-based brand equity: assets and liabilities linked to a brand, its name and symbol that add to or subtract from the value provided by a product or service to a firm

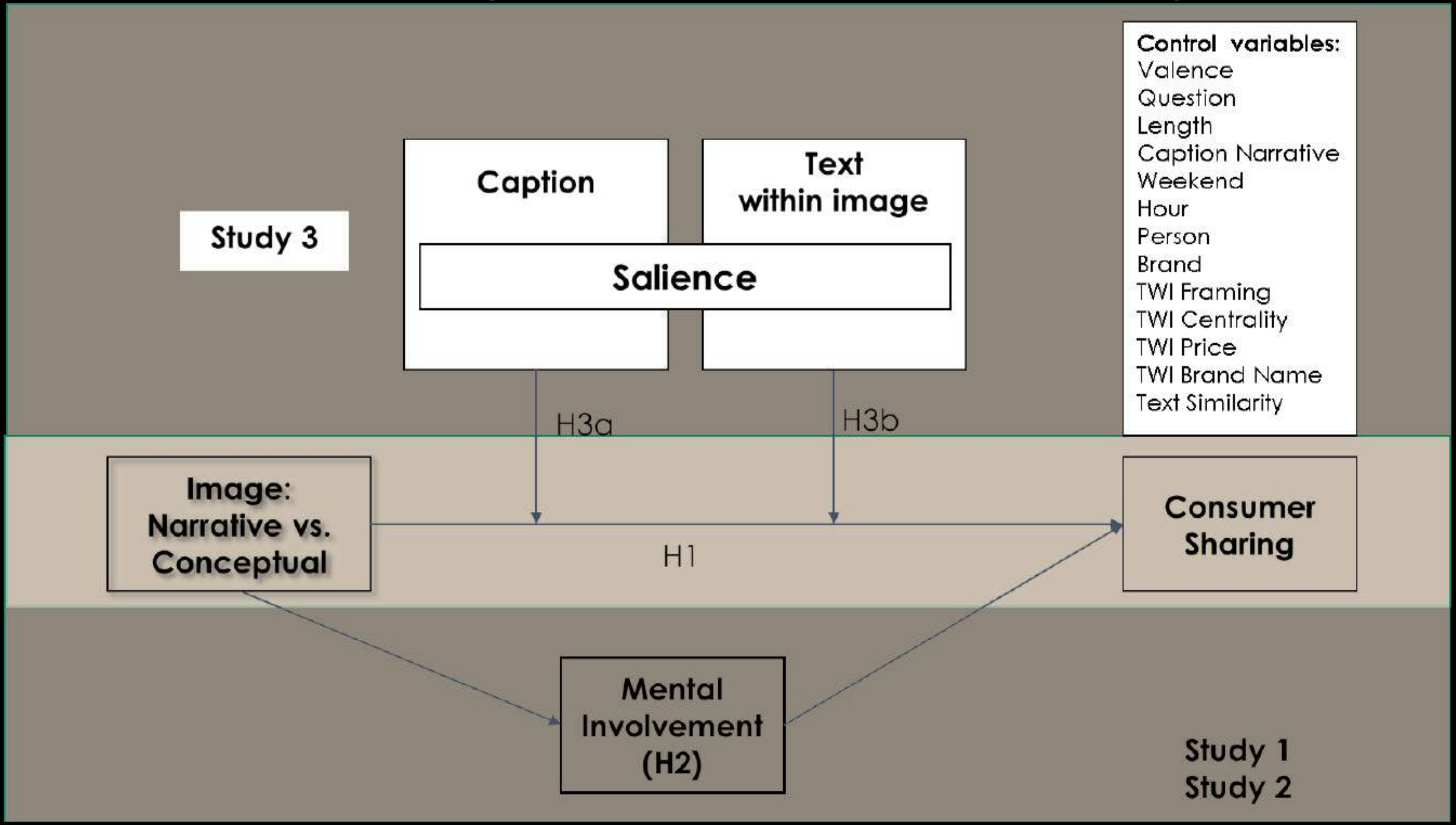
The Image + Text Formula in Social Media Brand Posts, Bridging Visual Semiotics and Consumer Sharing

Stefania Farace, Francisco Villarroel O., Dhruv Grewal and Ko de Ruyter

Deductive research result in “conceptual models” to be tested with research designs

When using deductive research we are generally assessing relationships of cause & effect between variables (Explanatory Models; Unit 4)

Fig. 1: Work in progress project about content marketing



References

- Batra, Rajeev, and Kevin Lane Keller. "Integrating marketing communications: New findings, new lessons, and new ideas." *Journal of Marketing* 80, no. 6 (2016): 122-145.
- Datta, Hannes, Kusum L. Ailawadi, and Harald J. Van Heerde. "How well does consumer-based brand equity align with sales-based brand equity and marketing-mix response?." Journal of Marketing 81, no. 3 (2017): 1-20.
- Hollebeek, Linda D., and Keith Macky. "Digital content marketing's role in fostering consumer engagement, trust, and value: Framework, fundamental propositions, and implications." Journal of Interactive Marketing 45 (2019): 27-41.
- Swaminathan, Vanitha, Alina Sorescu, Jan-Benedict EM Steenkamp, Thomas Clayton Gibson O'Guinn, and Bernd Schmitt. "Branding in a hyperconnected world: Refocusing theories and rethinking boundaries." Journal of Marketing 84, no. 2 (2020): 24-46.

Thanks

Unit 4 (Week 5): Content Marketing & Explanatory Analytics (Part 2)

Course: Business and Marketing Analytics

Francisco Villarroel Ordenes

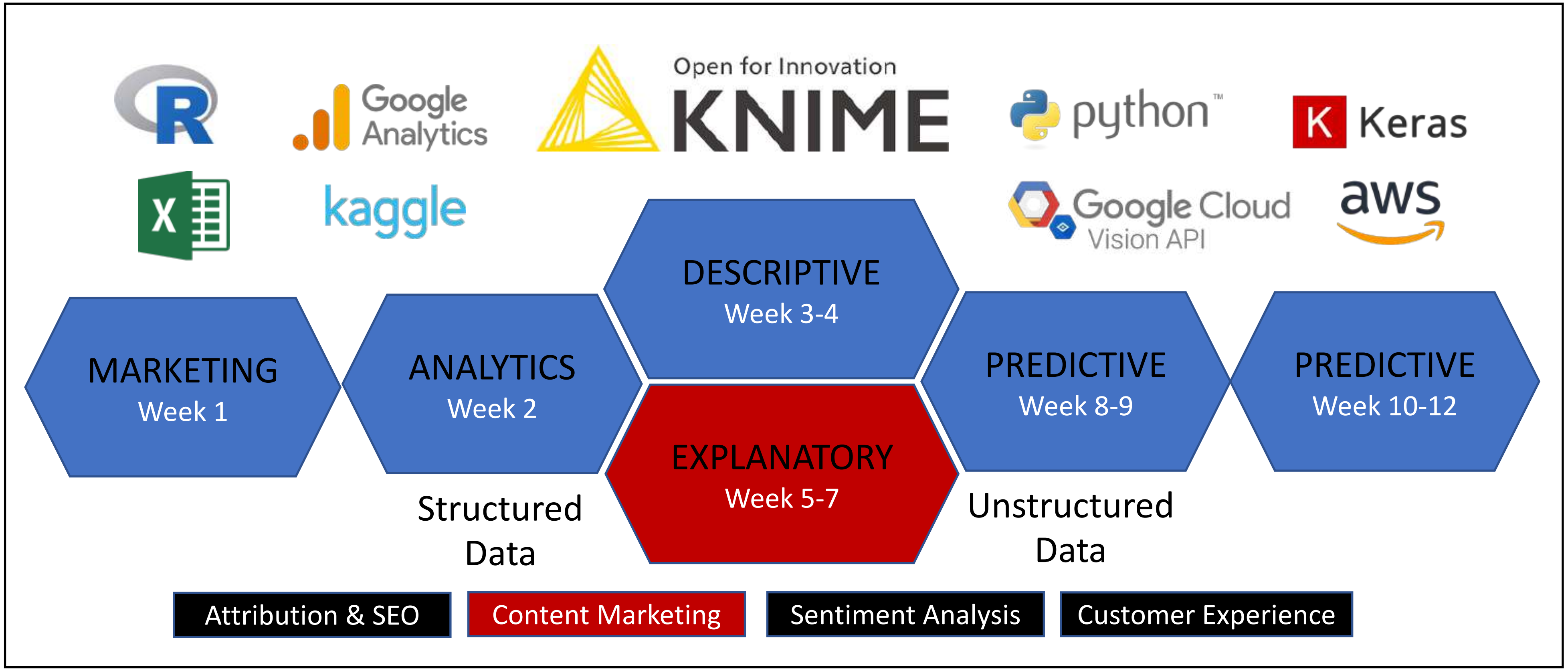
October, 2023

LUISS

Recap

Unit 4: Content Marketing and Explanatory Analytics

- Brand Equity
- Integrated Marketing Communication
- Content Marketing
- Social Media Marketing



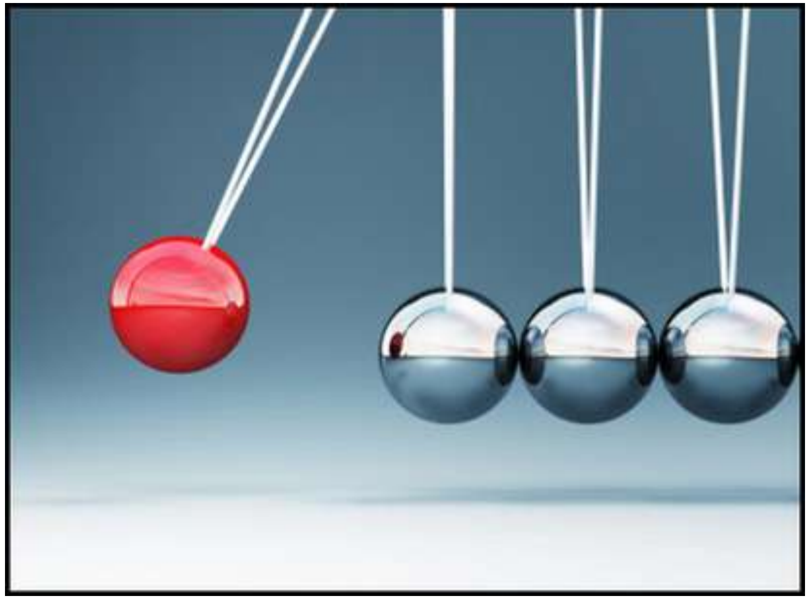
BUSINESS & MARKETING ANALYTICS FRAMEWORK

Objectives

- Applying explanatory analytics techniques in the field of social media content (this time not necessarily marketing).
- Introduction to conceptual models and hypothesis development (moderation & mediation)
- Introduction to regression models for count data (Negative Binomial)
- Results and Implications

Explanatory Analytics

- Diagnostic explanatory models that estimate relationships between variables and allow for hypothesis testing (Wedel and Kannan 2016)
- Other authors (Leeflang et al. 2020) use the term “descriptive models” to refer to explanatory analytics. The define descriptive models as models that describe demand and or supply relations on markets models that give answer to questions such as which marketing instruments affect sales or engagement



What causal relation(s) to test?

We should start from a relevant **managerial problem** at the consumer, firm and/or societal level.

How do we start testing relations?

When using **deductive reasoning** we start from the literature and identify a set of variables and relations that we are interested in testing

When using **inductive reasoning** we start from the data and identify patterns, constructs and variables. Then we might identify relations amongst them

Explanatory Analytics & Text Mining: News about Trump and their impact in online sharing

The Washington Post

The New York Times

USA TODAY

Formalized Models (Leeflang et. al 2020)

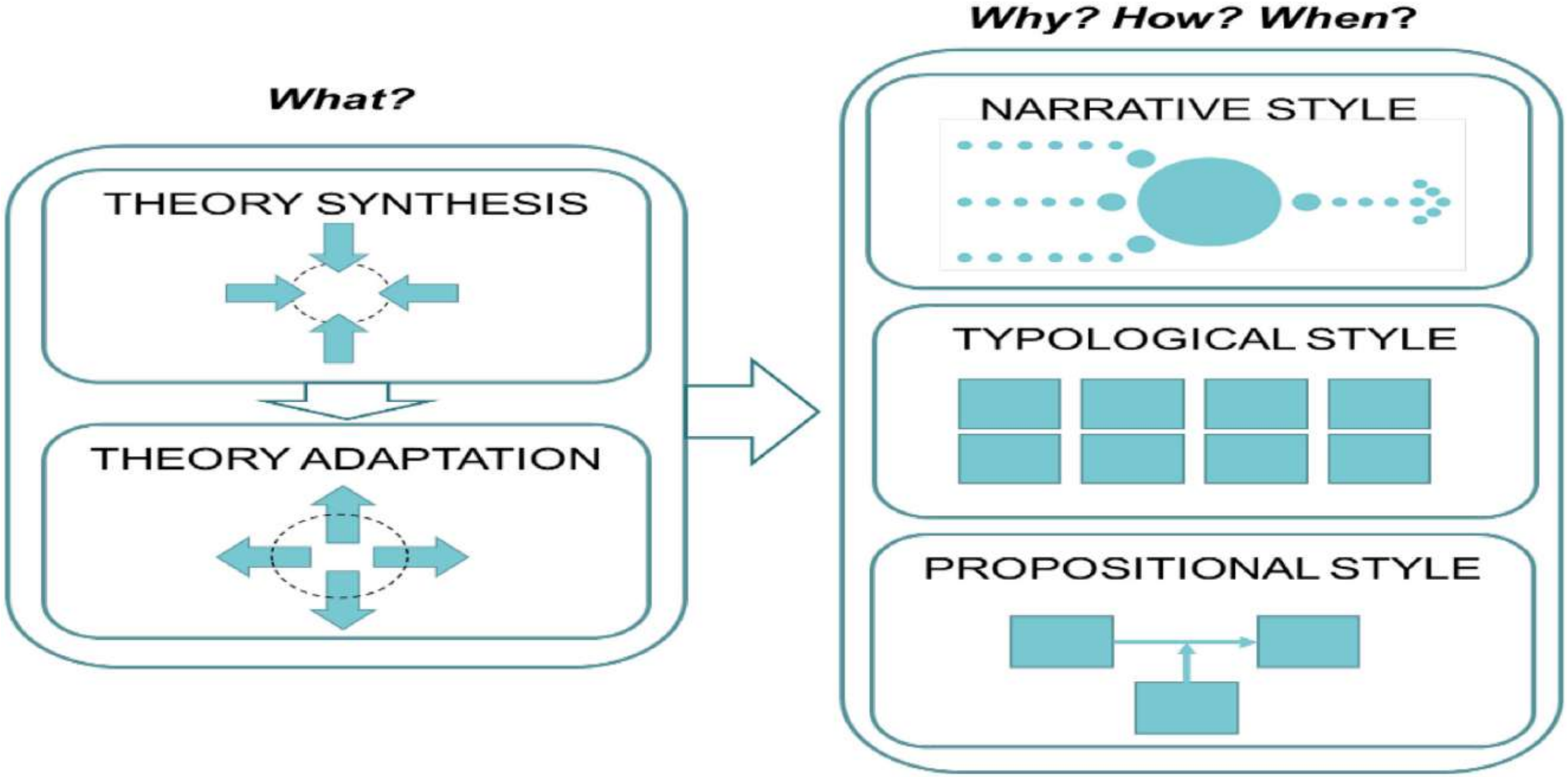
- Verbal Models: What a manager believes concerning marketing relations due to experience.
- Formalized Models
  - Diagram flows
  - Conceptual Frameworks
- Numerically Specified Models
  - Equation or set of equations that reflects marketing relations

Conceptual Frameworks and Theory

- The **way ideas are organized** to achieve a research project's purpose (Shields and Rangarajan 2013)
- **Informed by theory**, conceptual frameworks allow researchers to identify a sound set of relationships between managerial constructs of interest (customer satisfactions -> sales) (Lindgreen et al. 2020)
- A theory is a statement of **concepts and their interrelationships** that shows **when, how and/or why** a phenomenon occurs"
- The development of conceptual frameworks is **non-linear & iterative**

Types of Conceptual Framework

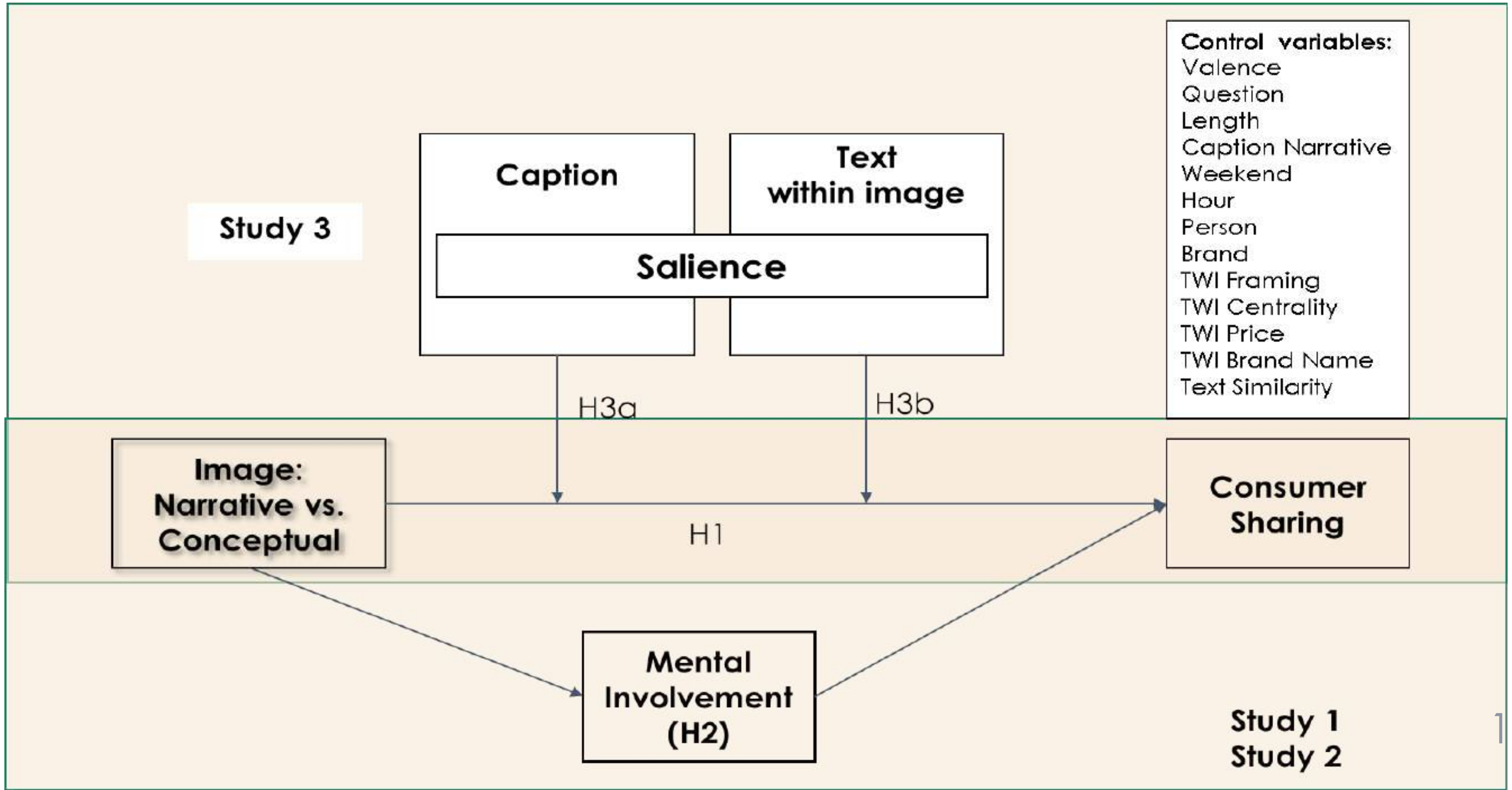
All types of conceptual framework aim answering why, how and when questions (Lindgreen et al. 2020)



Types of Conceptual Framework (1)

Proposition Based Style: The statement of theoretical propositions that introduces new constructs and cause-effect relationships

Farace et al. (2020): "The Text + Image Formula in Social Media Brand Posts that Really Works"

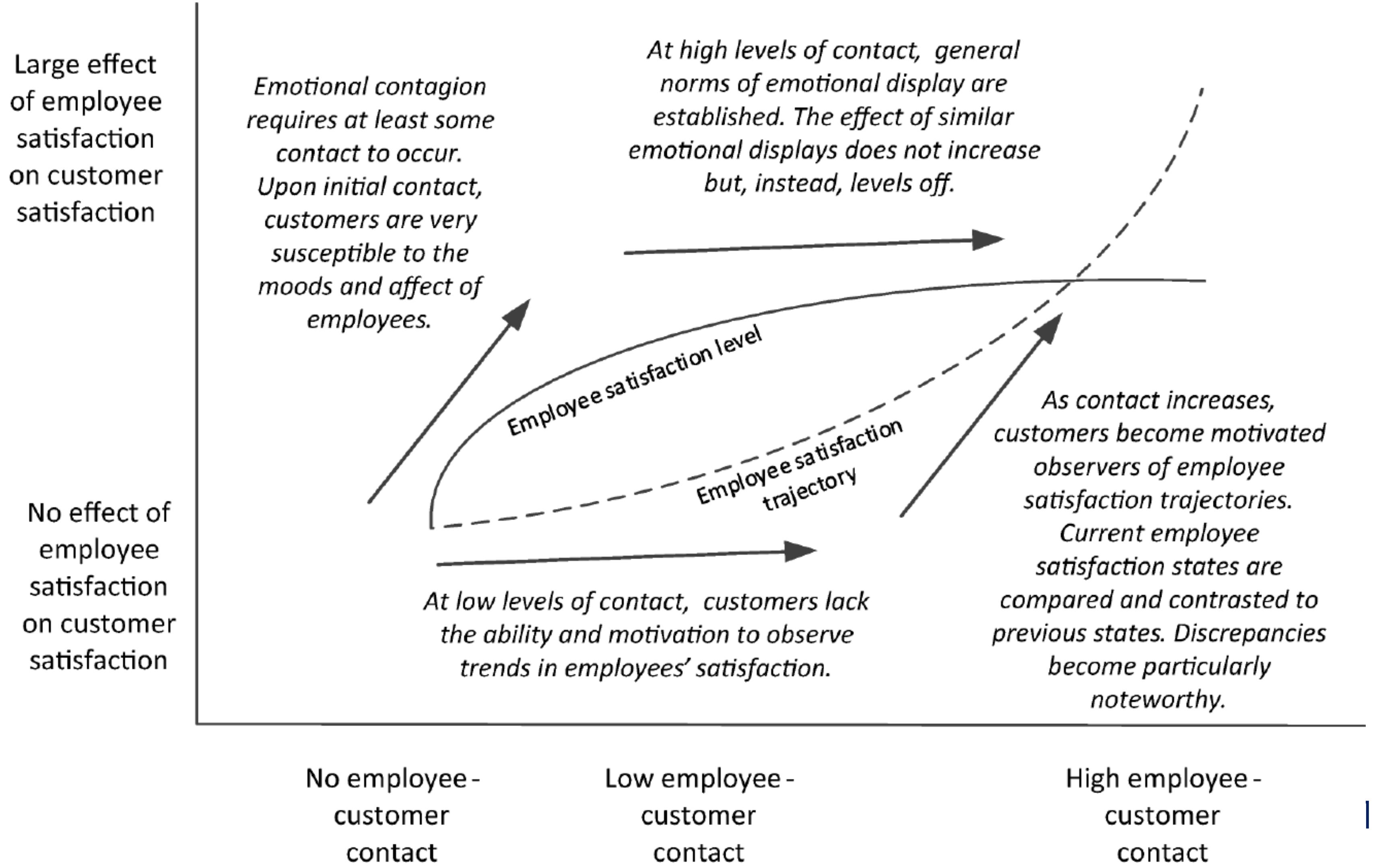


LUISS

Types of Conceptual Framework (2)

Narrative Based Style: The specification of a process model that lays out a set of mechanisms explaining events and outcomes.

Wolter et al. (2019): "Employee satisfaction trajectories & their effect on customer satisfaction & repatronage intentions"



LUISS

Types of Conceptual Framework (3)

Typology Based Style: The specification of a typology that interrelates different dimensions to flesh out new constructs and causal interactions.

Table 1. Evolution of Consumption: A Psychological Ownership Framework.

Dimension of Change	Threats to Psychological Ownership	Transfers of Psychological Ownership	Opportunities to Preserve Psychological Ownership
Legal ownership to legal access. Personal ownership of private goods is replaced with temporary access rights to use collectively consumed goods and services.	*Fractional ownership.* Bundle of rights associated with good divided among agents holding property rights to use, profit, change, or transfer ownership. ❖ Emphasize liquidity and economic value. *Impermanence.* Consumers no longer expect to keep goods—they assume goods will be returned, impairing reference-point shift to owner ("My…"). ❖ Extend/guarantee duration and consistency of consumption experience.	*Collective consumption.* Ownership felt for private goods transfers to goods used collectively ("MINE" to "OURS"). Reduced importance of individual goods, potential contaminated by dissociative group associations. Psychological ownership transfers to consumer communities. ❖ Develop object history/intimate knowledge, encourage self-investment, deploy counterconditioning, and develop consumer communities.	*More consumer choice.* Improved preference-matching due to more (often immediately) available options, increases perceived control. ❖ Provide larger assortments, increase mass customization. *New channels for self-expression.* Social media and reputation systems integral to access-based consumption platforms provide new outlets for social signaling. ❖ Develop social media applications and marketing strategy, encourage microblogging, offer access to aspirational brands/goods with positive signal value.

Types of Conceptual Framework (3)

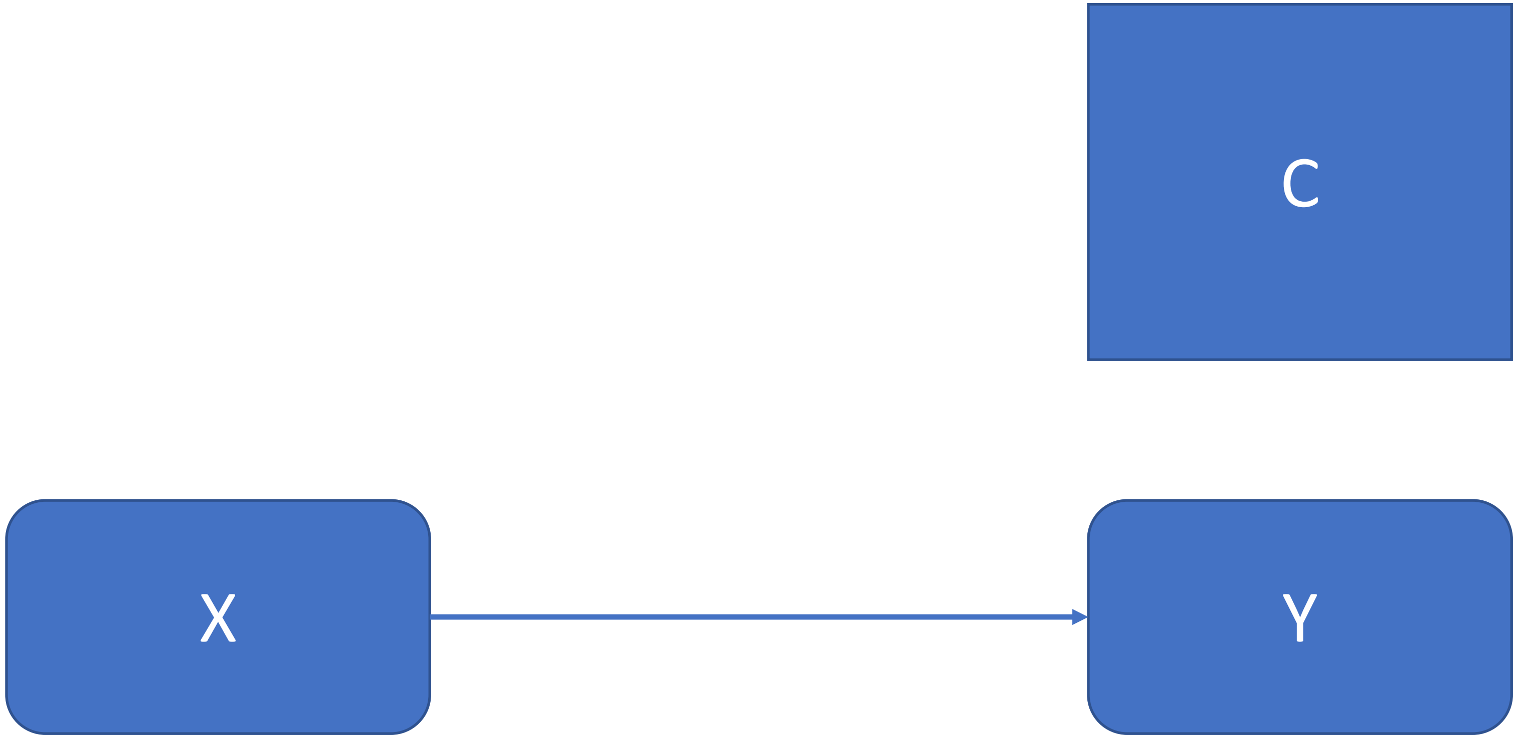
Table 1 (continuation; bottom part)

Material to experiential. Material goods are replaced with physical or digital experiential goods.	*Intangibility.* Consumers are less able to touch, hold, and physically manipulate experiential goods than physical goods. ❖ Develop haptic interfaces; interactive content; offer control over rate and timing of consumption; emphasize sensory features. *Reduced evaluability.* Ownership status is harder to determine (e.g., ownership of a vacation less clear than ownership of a vacation home). ❖ Make goods indexical connections–cues for personally meaningful events (e.g., cross sell physical goods, usage history reminders); gamification.	*Higher categorization level.* Category for which psychological ownership is experienced rises from individual goods to intermediary devices, platforms, and brands. ❖ Vertical integration, brand alliances, servitization, relationship marketing, intermediary device personalization.	*Greater self-identification.* Experiences are easier to integrate with self-concept than material goods (e.g., experiential purchases may generate more positive self-signals). ❖ Leverage identity marketing (e.g., "I am a skier" > "I own skis").

Notes: ❖ = recommended marketing actions to manage psychological ownership threats, transfers, and opportunities.

We will focus on proposition-based frameworks (ideal for explanatory analytics)

Main Effect (Proposition Based Style)



- Independent Variable (X)
- Dependent Variable (Y)
- Control Variables (C)

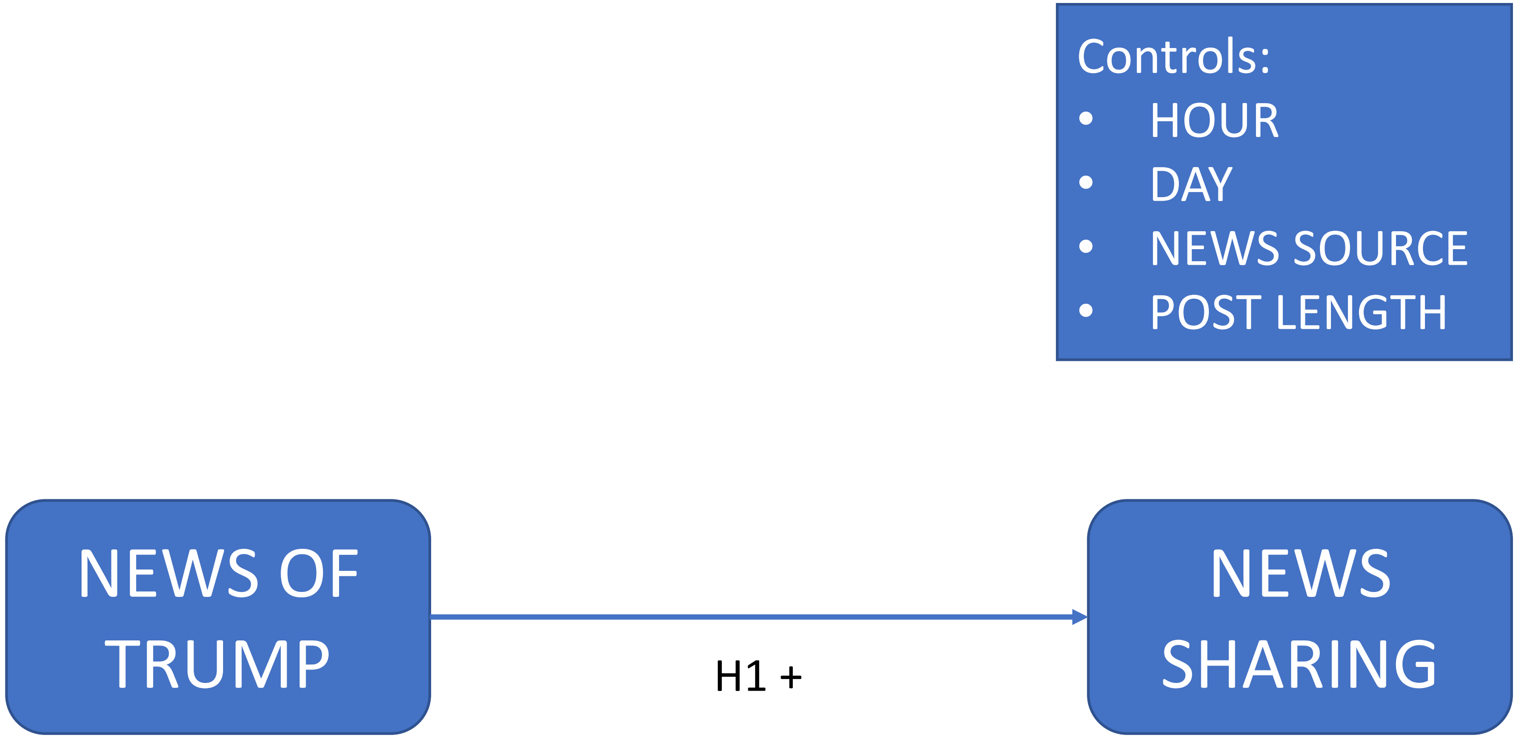
Main effect

- Imagine we want to test a hypothesis concerning news media containing information about Donald Trump and its effect on retweeting. How would you hypothesize?

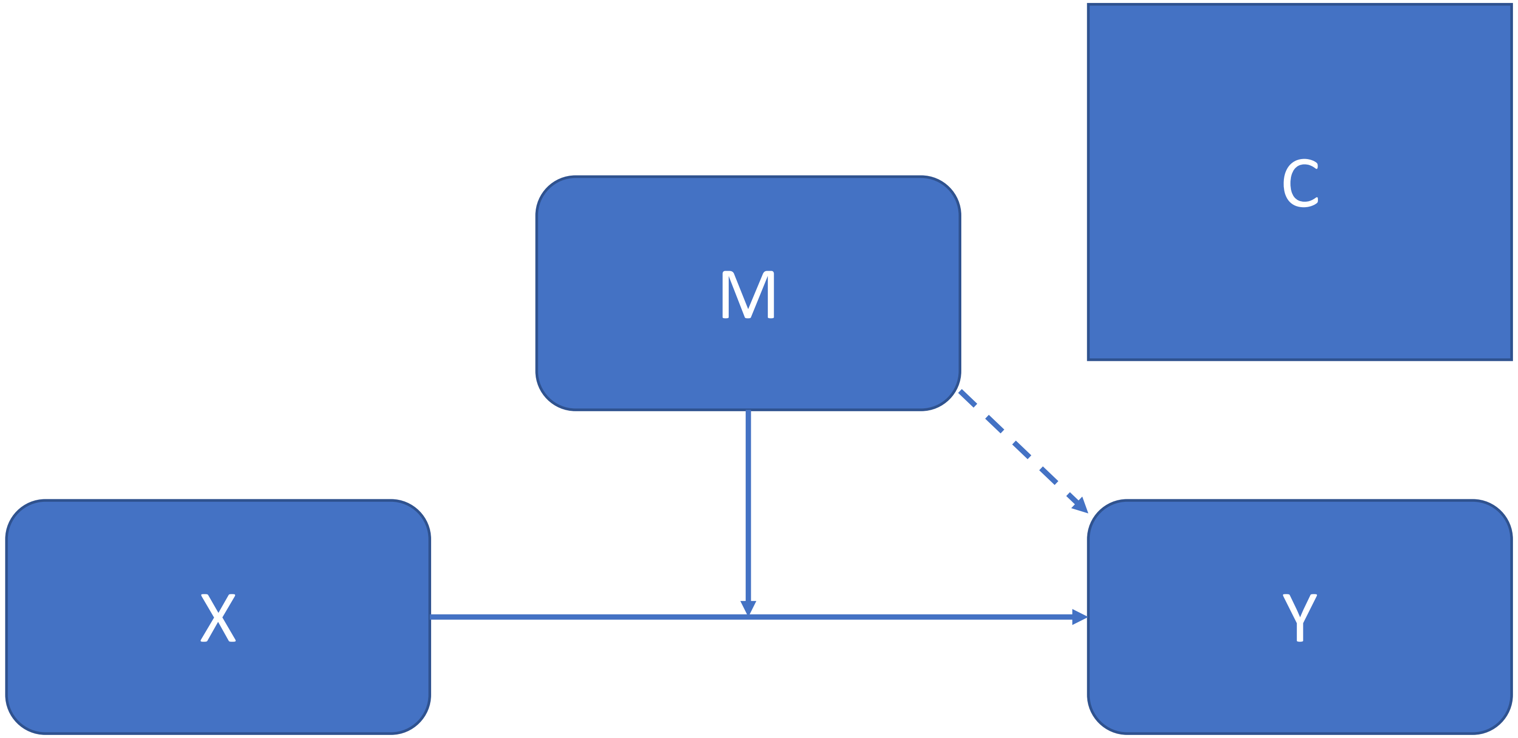
- We expect that news containing information about Donald Trump will be more relevant nowadays because of the US Election. In addition, because of his high presence on Twitter, news about Donald Trump should attract attention (Build up without theory)

H1 = News media tweets containing information about Donald Trump will be more shared than tweets non-related with Donald Trump

Main Effect


Controls:
• HOUR
• DAY
• NEWS SOURCE
• POST LENGTH
NEWS OF TRUMP
H1 +
NEWS SHARING

Moderation Effect



- Independent Variable (X)
- Dependent Variable (Y)
- Control Variables (C)
- Moderator Variable (M)

Moderator Effect

- A moderator is a variable that specifies conditions under which a given IV is related to a DV.
- Moderation implies an interaction effect, where introducing a moderating variable changes the direction or magnitude of the relationship between two variables.
- A moderation effect could be:
  - **Enhancing**, where increasing the moderator would increase or strengthen the effect of the IV on the DV
  - **Buffering**, where increasing the moderator would decrease or weaken the effect of the IV on the DV

Moderation effect

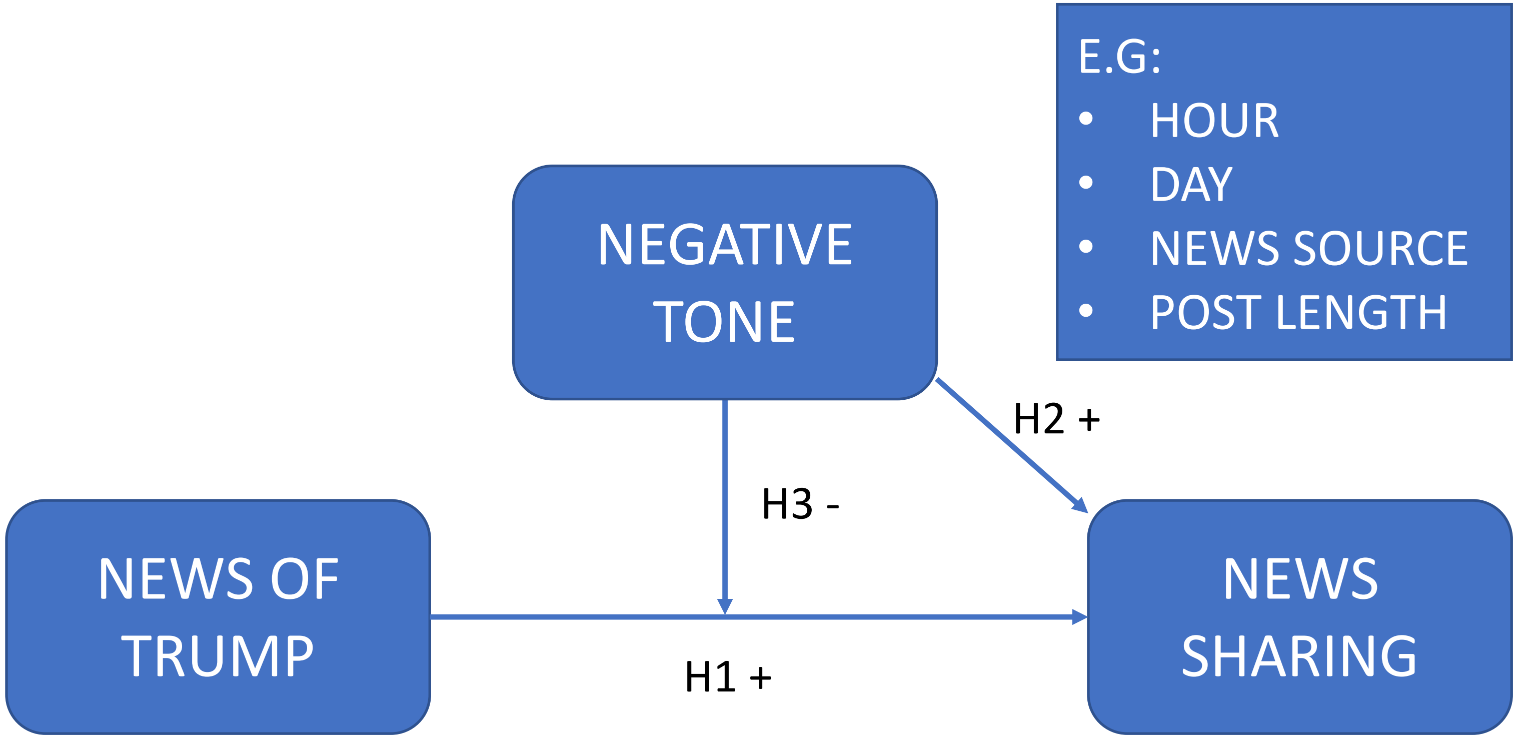
- What could be a moderator in our current model? Things that attenuate or strengthen sharing of news about Donald Trump
- **Negative Tone:** Drawing on "Negative Bias Theory" (things of more negative nature have more impact on people's behavior & cognition), we believe that greater negative news content will be more shared
- **TRUMP & Negative Tone:** The negative associations of TRUMP news, together with more negative news wording, will result in a double negative content. Excess of negative information will overload people with negativity, resulting in lesser sharing

H2 = Greater negative tone in news media tweets will result in greater consumer sharing

H3 = Greater negative tone in news media tweets will **weaken** the positive effect of TRUMP news on consumer sharing

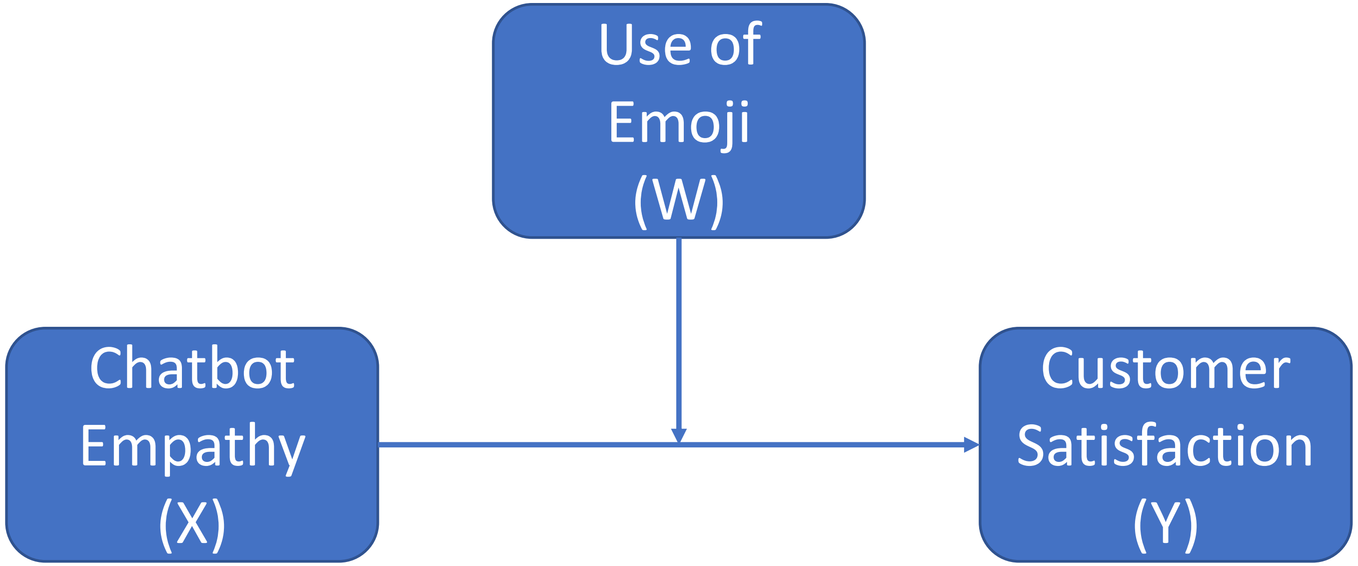
A Preview? “Tag cloud of Trump News”

Moderation Effect



Moderation Effect; Other Examples

- When (i.e., under what condition) is the relationship between chatbot empathy and customer satisfaction positive?
- For example, is chatbot use of emojis increasing the positive effect of chatbot empathy on perceived customer satisfaction?



- Independent Variable (X)
- Dependent Variable (Y)
- Moderator Variable (W)

A basic distinction (Leeflang et al. 2015)

Two types of Models in Marketing:

1. Models made for the primary purpose to support decision making of a specific marketing manager
   - Primary target audience are managers
   - It leads to case specific insights
   - Data might be focused on an industry or brand
2. Models that aim to advance general marketing knowledge
   - Primary target audience are marketing scientists/researchers
   - It leads to generalizable insights in marketing phenomena
   - Data should involve several brand/industries and periods

The Model Building Process (Leeflang et al. 2015)

1. Opportunity Identification: Managerial Problem.
2. Model Purpose: What problem should be solved?
3. Model Scope: What are the boundaries of it.
4. Data Availability: In this case Tweets!
5. Model Specification: Equation.
6. Estimation: OLS, ML, and others.
7. Validation: Make sure that it works (model fit, accuracy, etc.)
8. Cost-benefit consideration: Is it worth investing on it?
9. Use: How it can be used by a manager for decisions?
10. Updating: How to keep it working over time.

Model Criteria (Leeflang et al. 2015; Chapter 2)

- *Simple:* keeping the number of variables small, and only keep the relevant phenomena in the model
- *Evolutionary:* starting simple and expanding as time goes on (contributes to managers step-by-step understanding)
- *Complete:* account for all important variables. This will be relative to the problem, user and organization
- *Adaptive*: need to be updated regularly due changes in environment (e.g., parameters and structure when a new incumbent in the market, or changes in consumer base)
- *Robust*: a quality characteristic which makes it difficult for a user to obtain bad answers. The model builder needs to identify relevant constructs, define valid measures, specify meaningful functional forms (e.g., non-linearities), and accommodate appropriate interaction effects

How to test conceptual frameworks with Explanatory Analytics/Models

Several **regression models** can be used to test a conceptual model. The choice will depend on the assumptions of the model and generally the characteristics of the dependent variable.

- Linear (DV: Scale or Ratio)
- Logistic (DV: Yes or No; 1 or 0)
- Ordinal Logit (DV: Star Rating)
- Multinomial (DV: Categorical)
- Poisson & Negative Binomial Regression (DV: Count)
- Ridge/Lasso Regression (for multicollinearity)
- Bayesian Regression (probability based)

Linear Regression Models to Assess Causality

Correlations only provide insight into bi-variate relations, what we are looking for, is causal or predictive relations between independent and dependent variables in a model of the real world that allows us to control for as much noise as we can. For this we need to conduct a regression model.

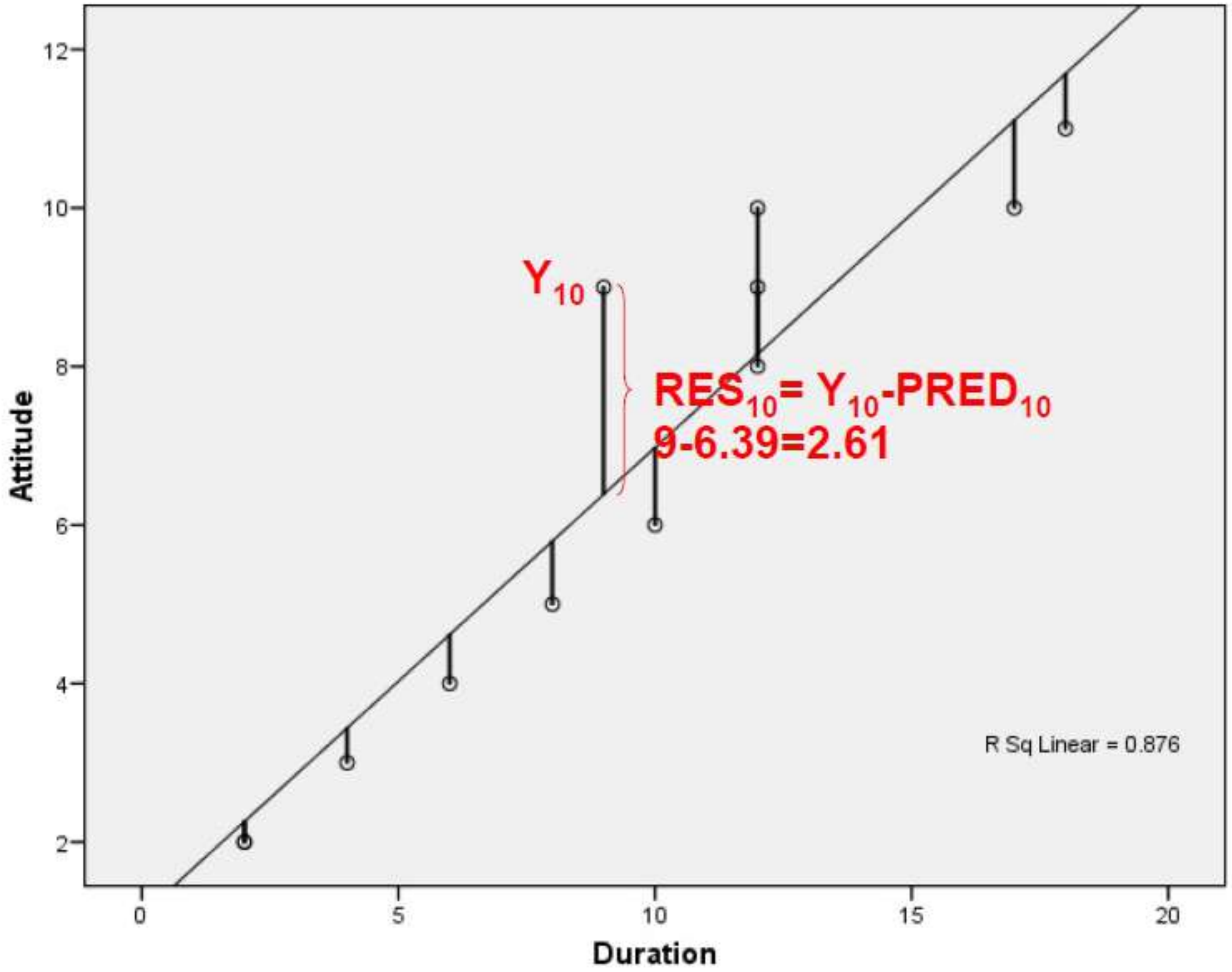
Model: $Y_i = \alpha + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \mu_i$

Estimation: *Ordinary Least Squares* (OLS; tailored for linear models) or *Maximum Likelihood* (ML; tailored for non-linear models). OLS aims to minimize the distance between Y and the predicted Y', ML is a likelihood maximization.

Fit Measures (how good your model is):

- *Adjusted* R^2 for OLS linear regression. Ranges from 0 (bad fit) to 1 (good fit
- *Log Likelihood, AIC and Pseudo R Square* for non linear models and ML

How it works (linear models)?



LUISS

Assumptions of Multiple Linear Regression (Wooldridge 2018; Chapter 3)

Under these assumptions parameters are unbiased:

1. **Linear in parameters.** There is a linear relationship between independent variables and the dependent variable. Yet, this is flexible because we can use different functions for dependent variable (e.g., log).
2. **Random sampling.** We have a random sample of n observations, and All Yi values are independent
3. **No perfect collinearity.** There does not exist any exact linear relationship between the Xi's (assumption of no multicollinearity). This can be assessed (preliminarily with a correlation table, or with VIF indicator)
4. **Zero conditional mean.** The ERROR "μ" has am expected value of zero given any values of the independent variables. It can fail if the relationship between explained and explanatory variables is misspecified (e.g., by forgetting to include a quadratic term). Omitting a variable that is correlated with any of the predictors would violate this assumption (**Endogeneity).**

***Homoskedasticity:** the error μ has the same variance given any values of the explanatory variables (if we include this assumption to the previous 4 we call them "Gauss-Markov Assumptions)

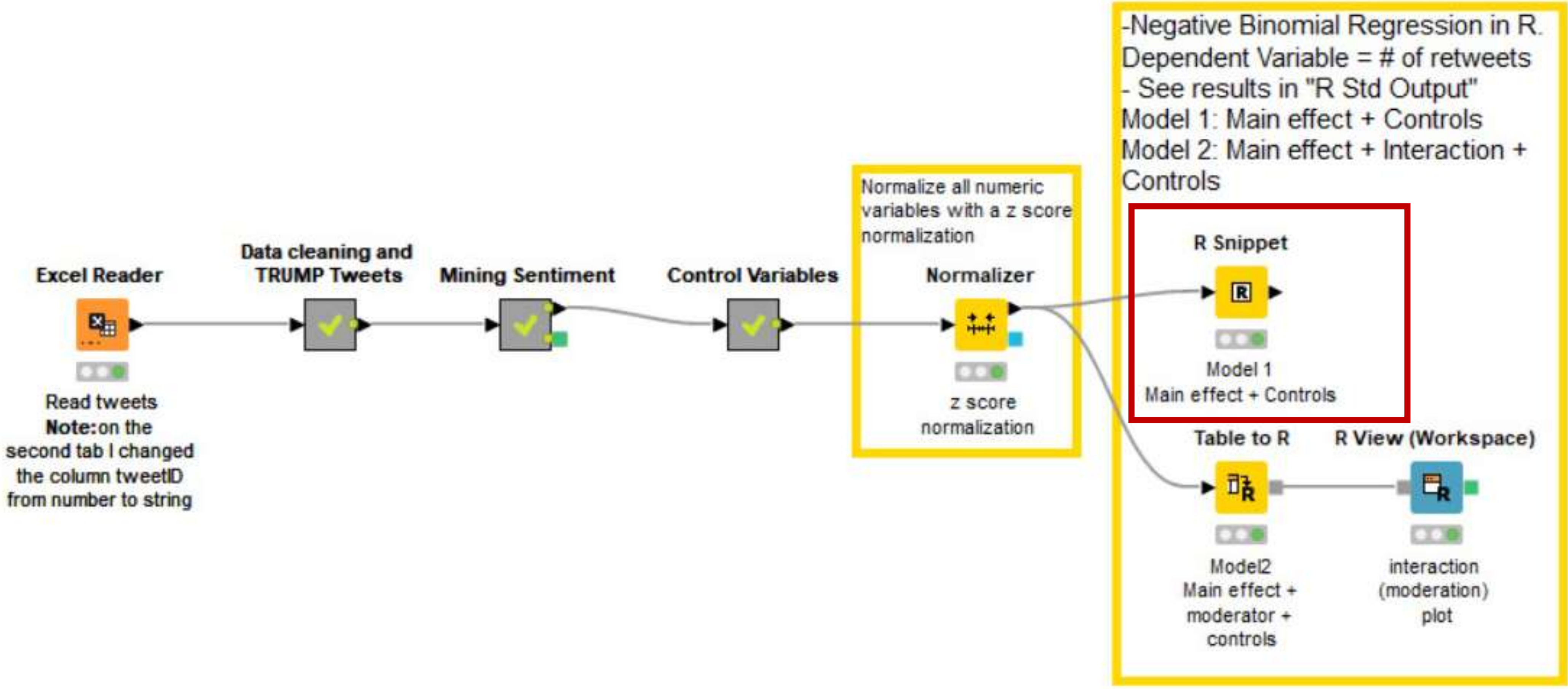
*Independent and dependent variables are metric (avoid nominal, counts; ratings can be used)

*These assumptions might vary when working with time series or panel data

Negative Binomial Regression

- Negative binomial regression is for modeling over dispersed (standard deviation greater than 2 times the mean) count variables.
- It is a non-linear type of model that uses maximum likelihood (ML; based on probability theory) for its estimation.
- Thus, the possible values of Y are the nonnegative integers: 0, 1, 2, 3, and so on.
- Negative binomial regression is a generalization of Poisson regression which loosens the restrictive assumption that the variance is equal to the mean made by the Poisson model

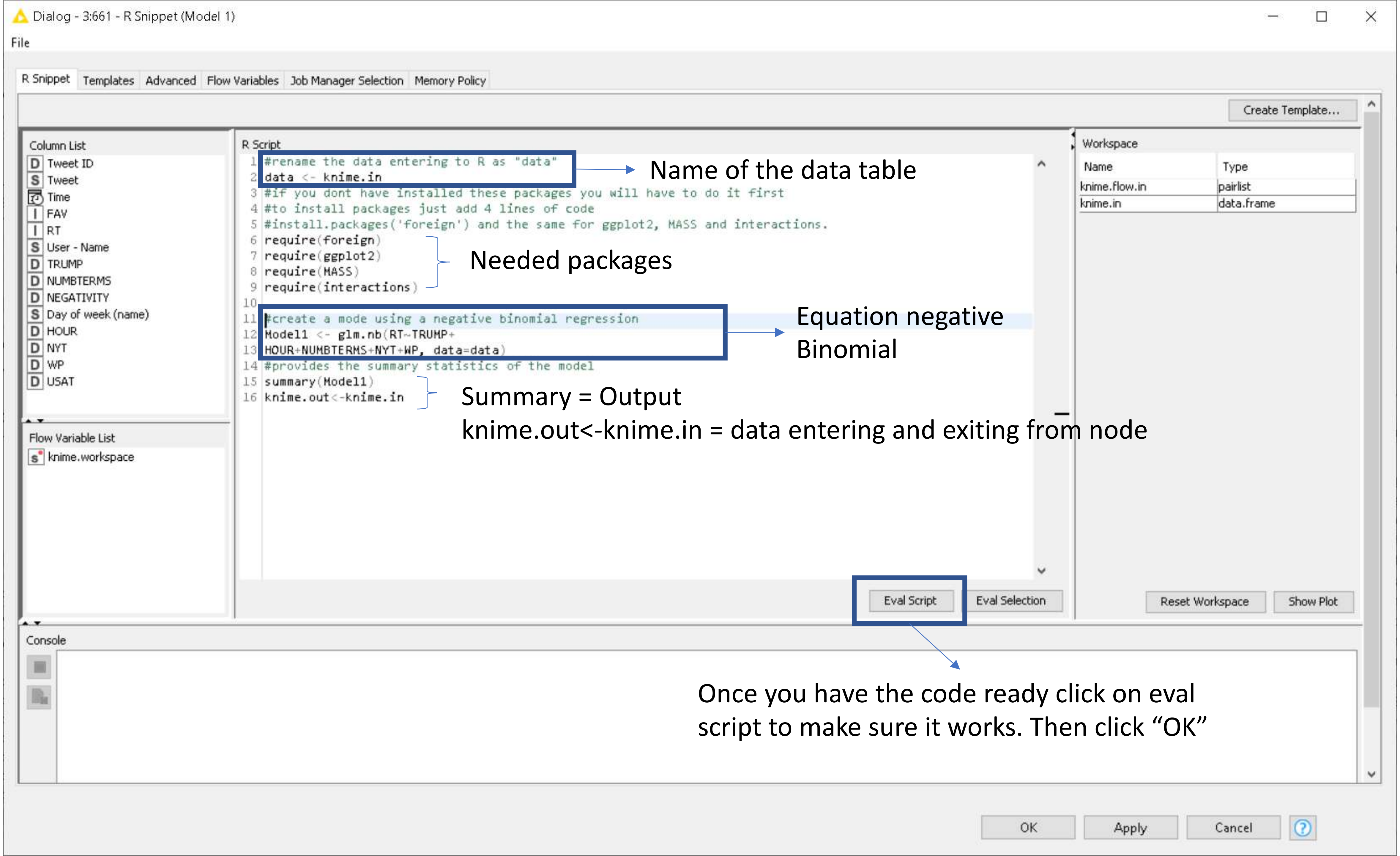
H1 = News media tweets containing information about Donald Trump will be more shared than tweets non related with Donald Trump



Model 1: Specification

$$Retweet_i = \exp(\propto + \beta_1 * TRUMP_i + \beta_2 * HOUR_i + \beta_3 * LENGTH_i + \beta_4 * NYT_i + \beta_5 * USATODAY_i + \mu_i)$$

- Dependent Variable = Retweet
- Main Effect (independent variable) = TRUMP
- Control Variables (to avoid omitted variables) = HOUR, LENGTH. Which others could be there?
- FIXED EFFECTS = NEWS MEDIA SOURCE (DUMMY VARIABLES). Which is the baseline of dummy variables?
- ERROR TERM = μ



Dialog - 3:661 - R Snippet (Model 1)
File
R Snippet
Templates
Advanced
Flow Variables
Job Manager Selection
Memory Policy
Create Template...
Column List
Tweet ID
Tweet
Time
FAV
RT
User - Name
TRUMP
NUMBTERMS
NEGATIVITY
Day of week (name)
HOUR
NYT
WP
USAT
Flow Variable List
knime.workspace
R Script
#rename the data entering to R as "data"
data <- knime.in
#if you dont have installed these packages you will have to do it first
#to install packages just add 4 lines of code
#install.packages('foreign') and the same for ggplot2, MASS and interactions.
require(foreign)
require(ggplot2)
require(MASS)
require(interactions)
#create a mode using a negative binomial regression
Model1 <- glm.nb(RT~TRUMP+
HOUR+NUMBTERMS+NYT+WP, data=data)
#provides the summary statistics of the model
summary(Model1)
knime.out<-knime.in
Name of the data table
Needed packages
Equation negative Binomial
Summary = Output
knime.out<-knime.in = data entering and exiting from node
Workspace
Name
Type
knime.flow.in
pairlist
knime.in
data.frame
Eval Script
Eval Selection
Reset Workspace
Show Plot
Console
Once you have the code ready click on eval script to make sure it works. Then click "OK"
OK
Apply
Cancel

Results Model 1(dependent variable: Retweets)

R Std Output - 3:661 - R Snippet (Model 1)

File

```
Call:
glm.nb(formula = RT ~ TRUMP + HOUR + NUMBTERMS + NYT + WP, data = data,
    init.theta = 0.6786637366, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8165  -1.0485  -0.6747  -0.1810  11.7599

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.79483    0.02522 190.117  < 2e-16 ***
TRUMP        0.24016    0.02548   9.427  < 2e-16 ***
HOUR        -0.19800    0.02550  -7.766 8.13e-15 ***
NUMBTERMS   -0.09451    0.03411  -2.771   0.0056 **
NYT          0.98638    0.03091  31.907  < 2e-16 ***
WP           0.74594    0.03412  21.863  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6787) family taken to be 1)

    Null deviance: 3837.7  on 2336  degrees of freedom
Residual deviance: 2801.0  on 2331  degrees of freedom
AIC: 26868

Number of Fisher Scoring iterations: 1


              Theta:  0.6787
          Std. Err.:  0.0171

 2 x log-likelihood:  -26853.7500
```

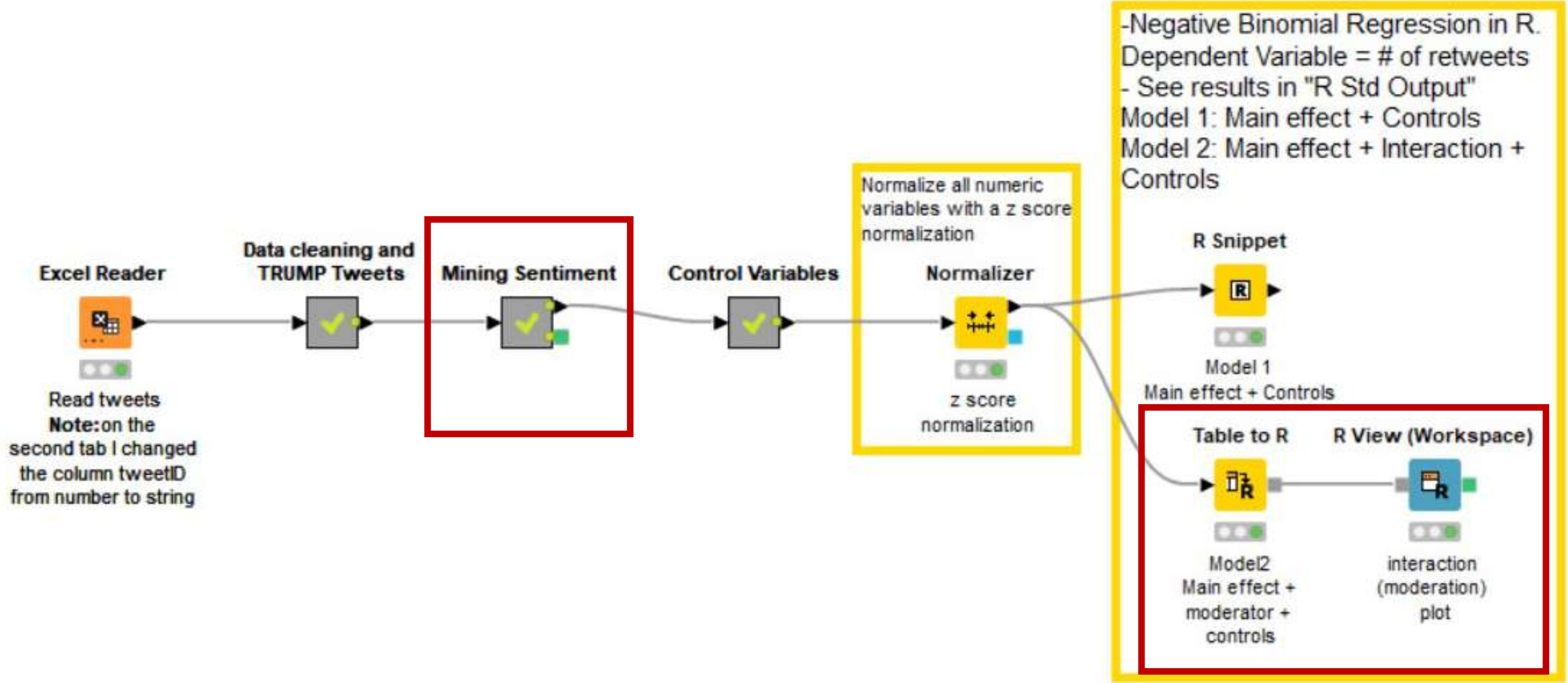

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.79483    0.02522 190.117  < 2e-16 ***
TRUMP        0.24016    0.02548   9.427  < 2e-16 ***
HOUR        -0.19800    0.02550  -7.766 8.13e-15 ***
NUMBTERMS   -0.09451    0.03411  -2.771   0.0056 **
NYT          0.98638    0.03091  31.907  < 2e-16 ***
WP           0.74594    0.03412  21.863  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6787) family taken to be 1)

    Null deviance: 3837.7  on 2336  degrees of freedom
Residual deviance: 2801.0  on 2331  degrees of freedom
AIC: 26868
```


*Standard for statistical significance: P value <0.05

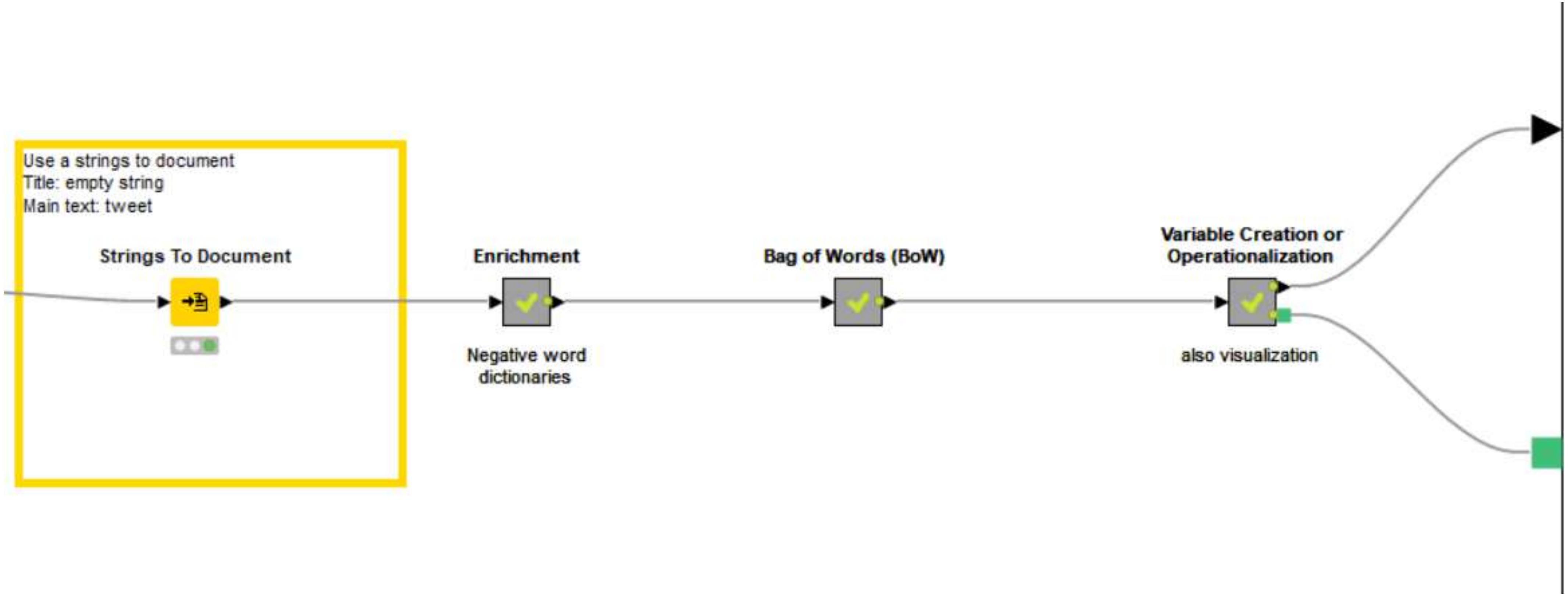
Interpretation?

H2 = Greater negative tone in news media tweets will result in greater consumer sharing

H3 = Greater negative tone in news media tweets will **weaken** the negative effect of TRUMP news on consumer sharing



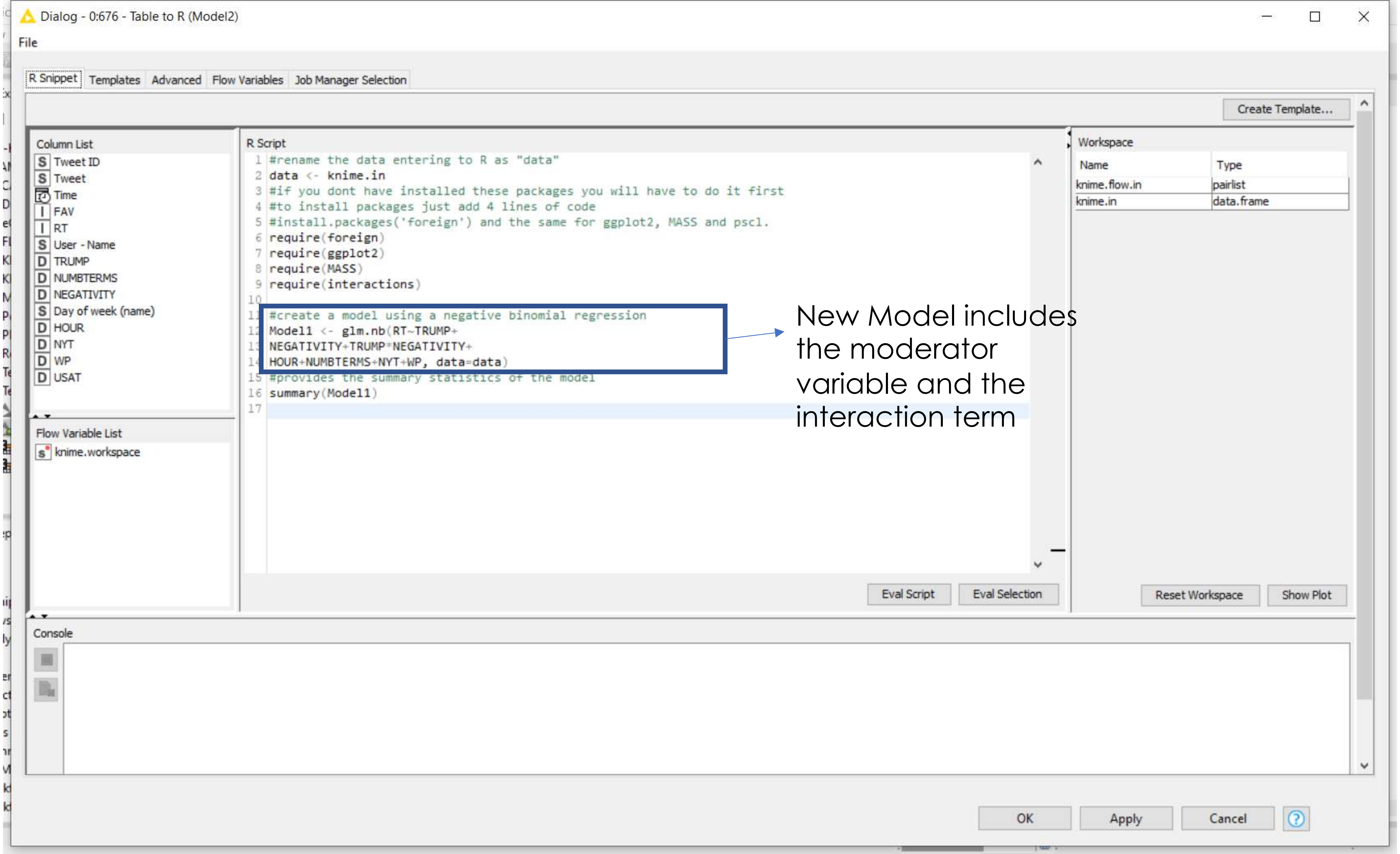
Metanode: Mining Sentiment (Negativity) Dictionary Method for Variable Operationalization



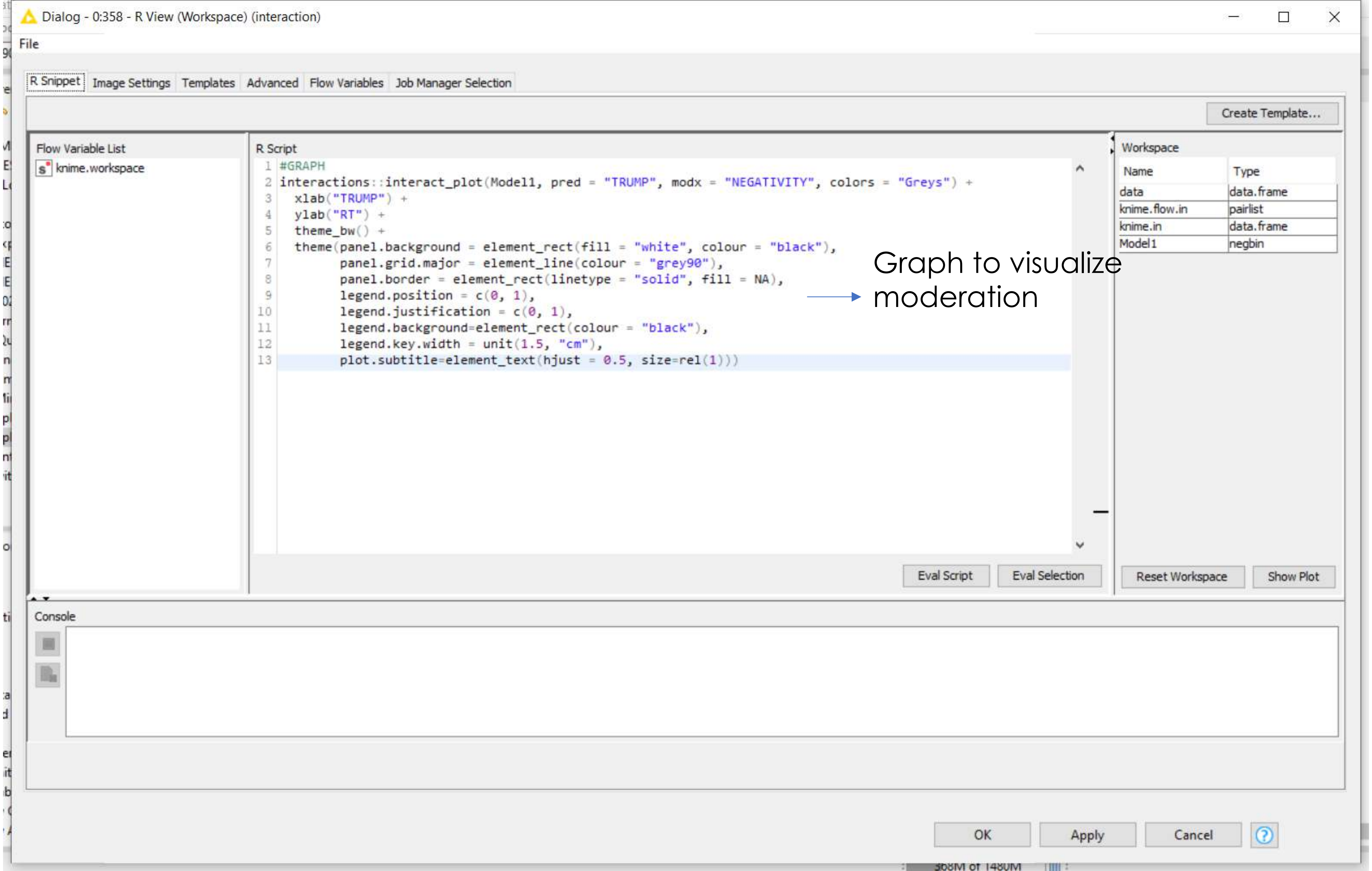
Model 2: Specification

$$Retweet_i = \exp\left(\propto + \beta_1 * TRUMP_i + \underbrace{\boldsymbol{\beta_2 * NEGATIVITY_i}}_{\text{Moderator Variable}} + \boldsymbol{\beta_3} * \underbrace{\boldsymbol{TRUMP_i * NEGATIVITY_i}}_{\text{Interaction Term}} + \beta_4 * HOUR_i + \beta_5 * LENGTH_i + \beta_6 * NYT_i + \beta_7 * USATODAY_i + \mu_i\right)$$

- Dependent Variable = Retweet
- Main Effect (independent variable) = TRUMP
- **Moderator/Interaction: NEGATIVITY**
- Control Variables (to avoid omitted variables) = HOUR, LENGTH. Which others could be there?
- FIXED EFFECTS = NEWS MEDIA SOURCE (DUMMY VARIABLES). Which is the baseline of dummy variables?
- ERROR TERM = μ



Dialog - 0:676 - Table to R (Model2)
File
R Snippet
Templates
Advanced
Flow Variables
Job Manager Selection
Create Template...
Column List
Tweet ID
Tweet
Time
FAV
RT
User - Name
TRUMP
NUMBTERMS
NEGATIVITY
Day of week (name)
HOUR
NYT
WP
USAT
Flow Variable List
knime.workspace
R Script
#rename the data entering to R as "data"
data <- knime.in
#if you dont have installed these packages you will have to do it first
#to install packages just add 4 lines of code
#install.packages('foreign') and the same for ggplot2, MASS and pscl.
require(foreign)
require(ggplot2)
require(MASS)
require(interactions)
#create a model using a negative binomial regression
Model1 <- glm.nb(RT~TRUMP+
NEGATIVITY+TRUMP*NEGATIVITY+
HOUR+NUMBTERMS+NYT+WP, data=data)
#provides the summary statistics of the model
summary(Model1)
Workspace
Name
Type
knime.flow.in
pairlist
knime.in
data.frame
Eval Script
Eval Selection
Reset Workspace
Show Plot
Console
OK
Apply
Cancel
New Model includes the moderator variable and the interaction term



Graph to visualize moderation

Results Model 2(dependent variable: Retweets)

R Std Output - 3:669 - R Snippet (Model2)

File

```
Call:
glm.nb(formula = RT ~ TRUMP + NEGATIVITY + TRUMP * NEGATIVITY +
    HOUR + NUMBTERMS + NYT + WP, data = data, init.theta = 0.6827072274,
    link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8187  -1.0427  -0.6810  -0.1781  12.5455

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        4.78678    0.02516 190.286  < 2e-16 ***
TRUMP              0.24079    0.02543   9.469  < 2e-16 ***
NEGATIVITY         0.04623    0.02576   1.795 0.072692 .
HOUR              -0.18257    0.02543  -7.178 7.06e-13 ***
NUMBTERMS         -0.10051    0.03401  -2.955 0.003127 **
NYT                0.97975    0.03083  31.784  < 2e-16 ***
WP                 0.72041    0.03420  21.062  < 2e-16 ***
TRUMP:NEGATIVITY -0.09310    0.02828  -3.292 0.000995 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6827) family taken to be 1)

    Null deviance: 3860.2  on 2336  degrees of freedom
Residual deviance: 2798.8  on 2329  degrees of freedom
AIC: 26853

Number of Fisher Scoring iterations: 1


              Theta:  0.6827
          Std. Err.:  0.0172

 2 x log-likelihood:  -26835.1780
```

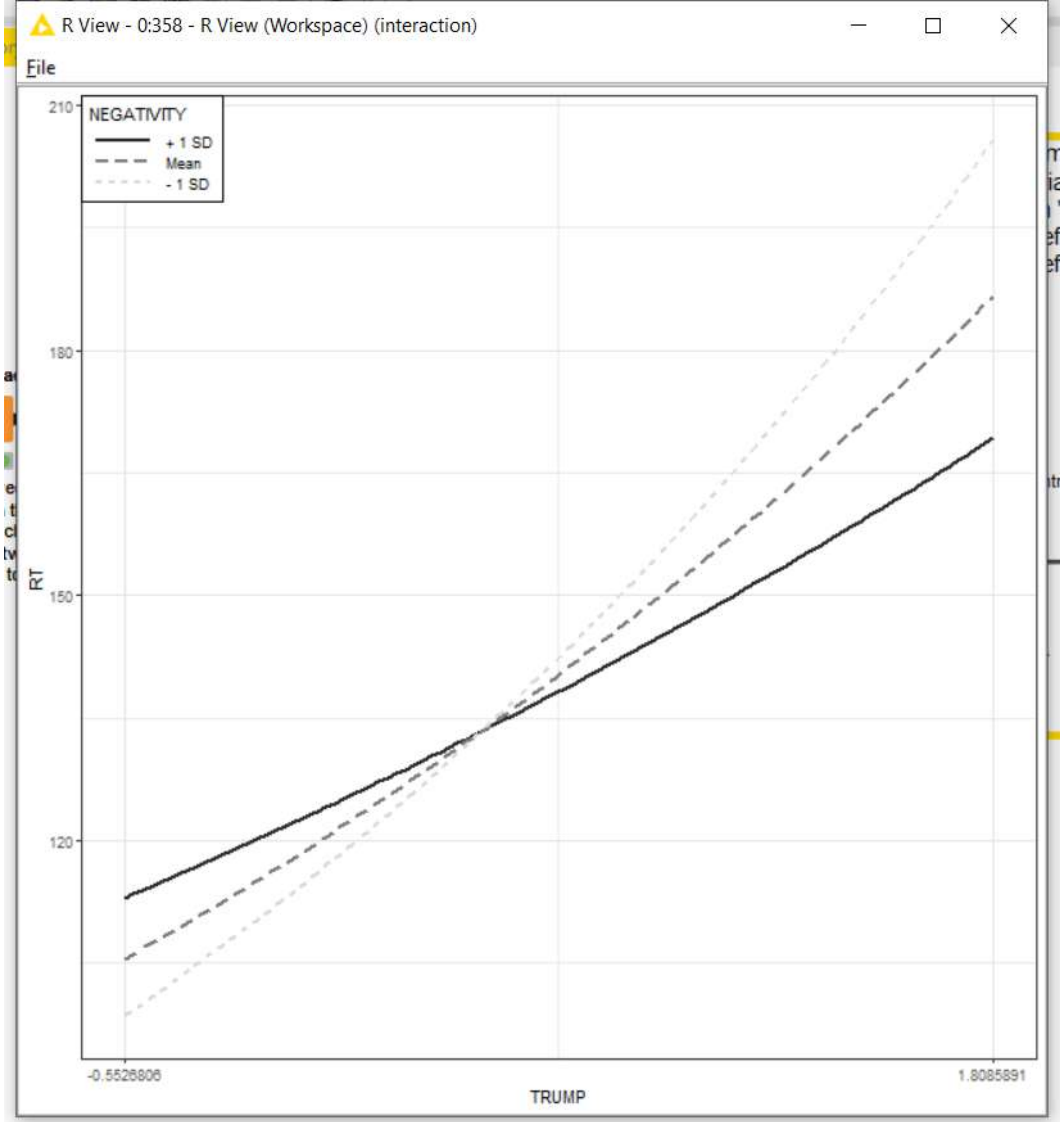
```
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        4.79233    0.02519 190.255  < 2e-16 ***
TRUMP              0.24151    0.02544   9.492  < 2e-16 ***
NEGATIVITY         0.02935    0.02542   1.154  0.24832
HOUR              -0.19077    0.02548  -7.488    7e-14 ***
NUMBTERMS         -0.10375    0.03411  -3.041  0.00236 **
NYT                0.98518    0.03088  31.902  < 2e-16 ***
WP                 0.73597    0.03427  21.473  < 2e-16 ***
TRUMP:NEGATIVITY -0.07035    0.02637  -2.668  0.00763 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6804) family taken to be 1)

    Null deviance: 3847.3  on 2336  degrees of freedom
Residual deviance: 2800.0  on 2329  degrees of freedom
AIC: 26864
```

*Standard for statistical significance: P value <0.05

Interpretation?

Interaction Plot

- Interpretation?



Model Comparison. DV = Retweets

Coefficients	MODEL 1	MODEL 2
(Intercept)	4.794**	4.792**
TRUMP	0.240**	0.241**
NEGATIVITY		0.029
HOUR	-0.198**	-0.190**
NUMBTERMS	-0.094**	-0.103**
NYT	0.986**	0.985**
WP	0.745**	0.735**
TRUMP:NEGATIVITY		-0.070**
AIC	26868	26864
N	2337	2337

- H1 = Supported
- H2 = Not Supported
- H3 = Supported
- Model Fit ?
- Improved from model 1 -> 2

Conclusions

The example demonstrated the most important steps in the explanatory analytics process, combining 1)managerial problem, 2)conceptualization and theory, 3)measurement and 4)model development.

1. Conceptualization
2. Hypothesis & theory
3. Model development
4. Measurement (Text Mining)
5. Estimation (Negative Binomial Regression
6. Hypothesis Testing
7. Findings & Conclusion

Beyond this simple exercise

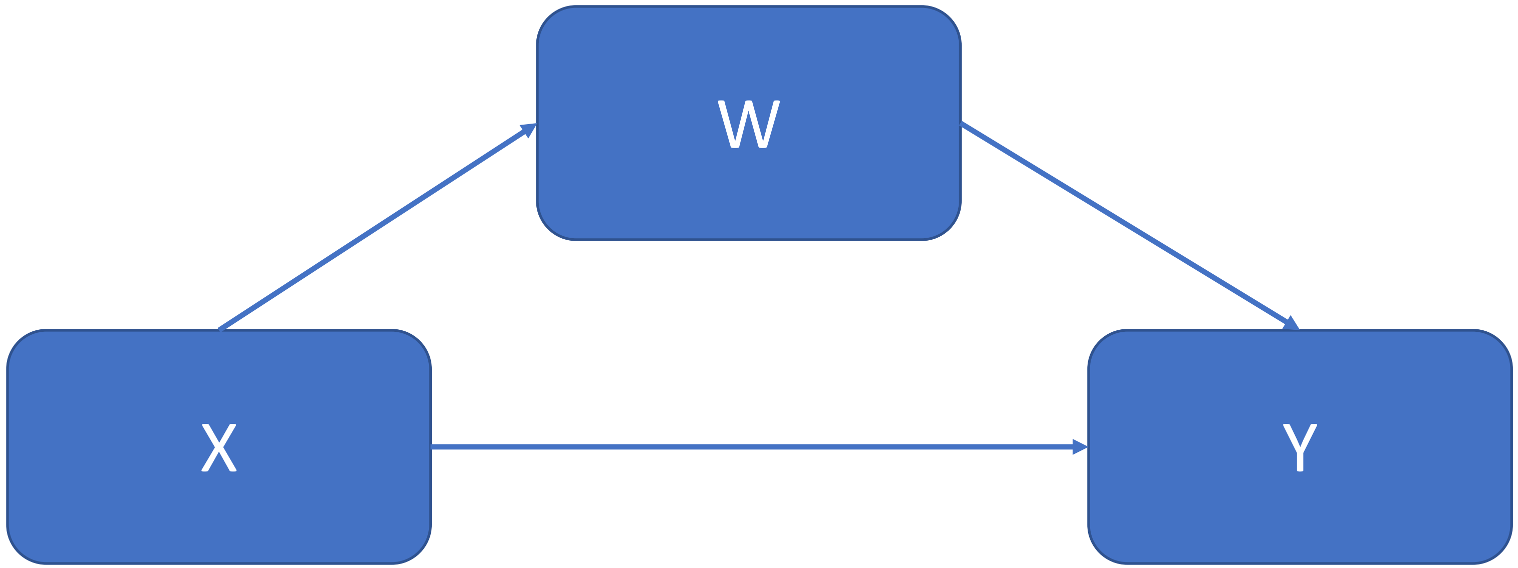
- Endogeneity Concerns
  - Selection Bias
  - Dynamic models (control functions)
- Panel Data Analysis & Time Series
- Other types of regression

References

- Book: Leeflang P., Wieringa J., Bijmolt T., and Pauwels K. (2015), “Modeling Markets, Analyzing Marketing Phenomena and Improving Marketing Decision Making.” Springer, NY.
- Lindgreen, Adam, C. Anthony Di Benedetto, Roderick J. Brodie, and Elina Jaakkola. "How to develop great conceptual frameworks for business-to-business marketing." (2020).
- Shields, Patricia M., and Nandhini Rangarajan. A playbook for research methods: Integrating conceptual frameworks and project management. New Forums Press, 2013.
- Wooldridge, Jeffrey M. Introductory econometrics: A modern approach. Nelson Education, 2016.

Thanks

Mediation Effect



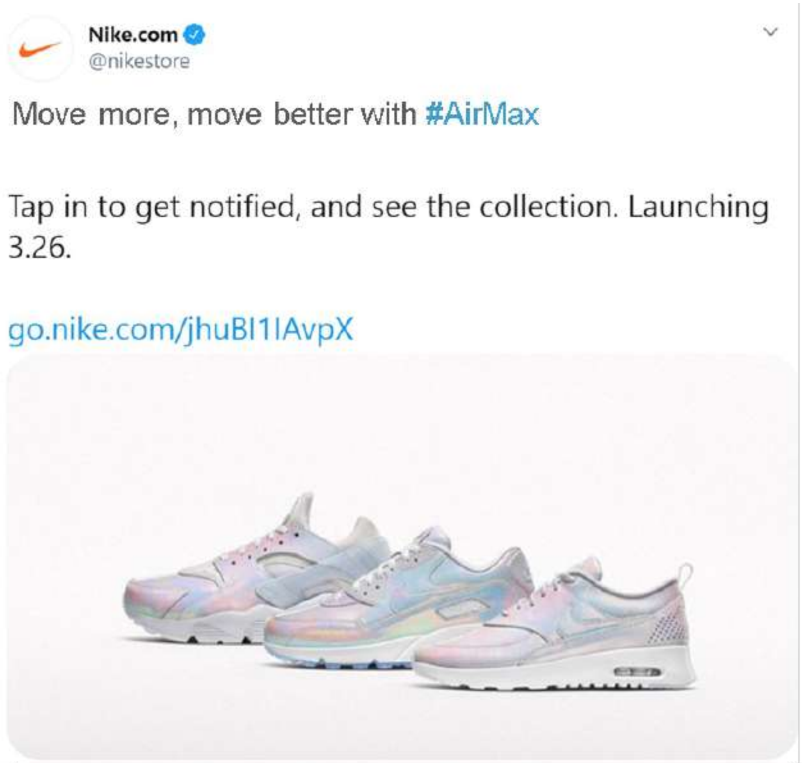
- Independent Variable (X)
- Dependent Variable (Y)
- Mediator Variable (M)

Mediation Effect

- Mediation implies a situation where the effect of the independent variable on the dependent variable can best be explained using a third mediator variable which is caused by the independent variable and is itself a cause for the dependent variable
- The mediator explains "how" (or some people says "why") a DV and IV are related
- A mediator is a variable that intervenes/intermediates in the relation between an independent variable and an outcome
- That is to say instead of X causing Y directly, X is causing the mediator M, and M is in turn causing Y. The causal relationship between X and Y in this case is said to be indirect.

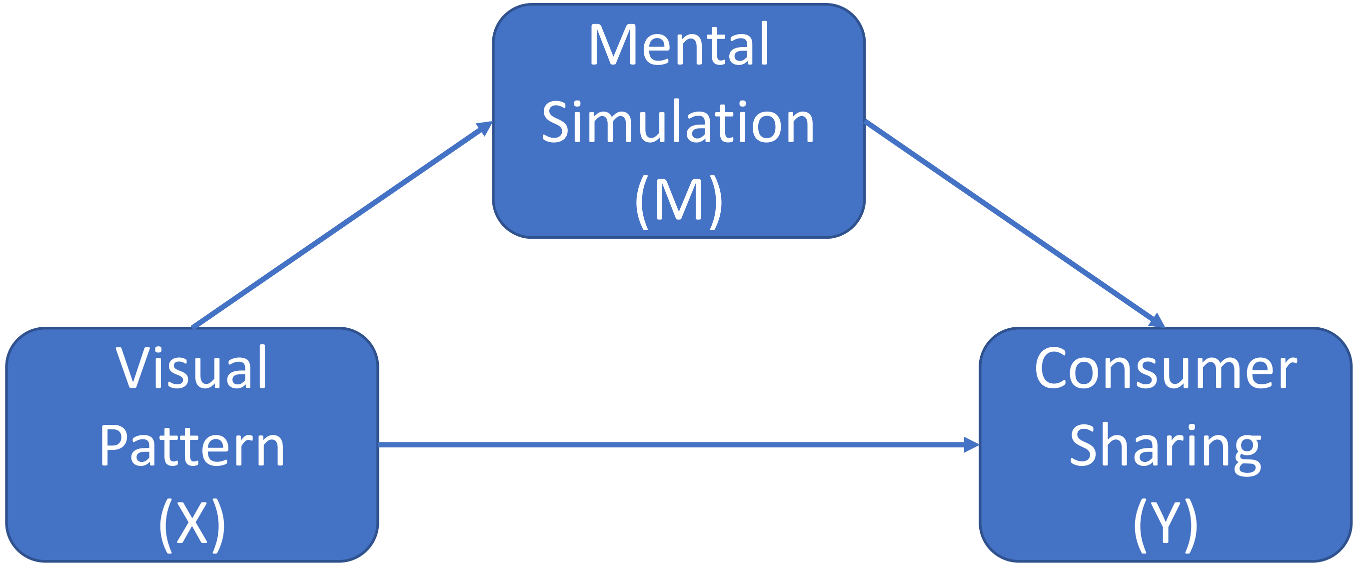
Mediation Effect

- How or why do regular visual patterns enhance consumer sharing in social media?
- For example, does the level of mental simulation resulting from the visual patter affect the consumer sharing?



Mediation Effect

- How or why do regular visual patterns enhance consumer sharing in social media?
- For example, does the level of mental simulation resulting from the visual patter affect the consumer sharing?



- Independent Variable (X)
- Dependent Variable (Y)
- Moderator Variable (W)

Google Keyword Planner Process

Course: Business and Marketing Analytics

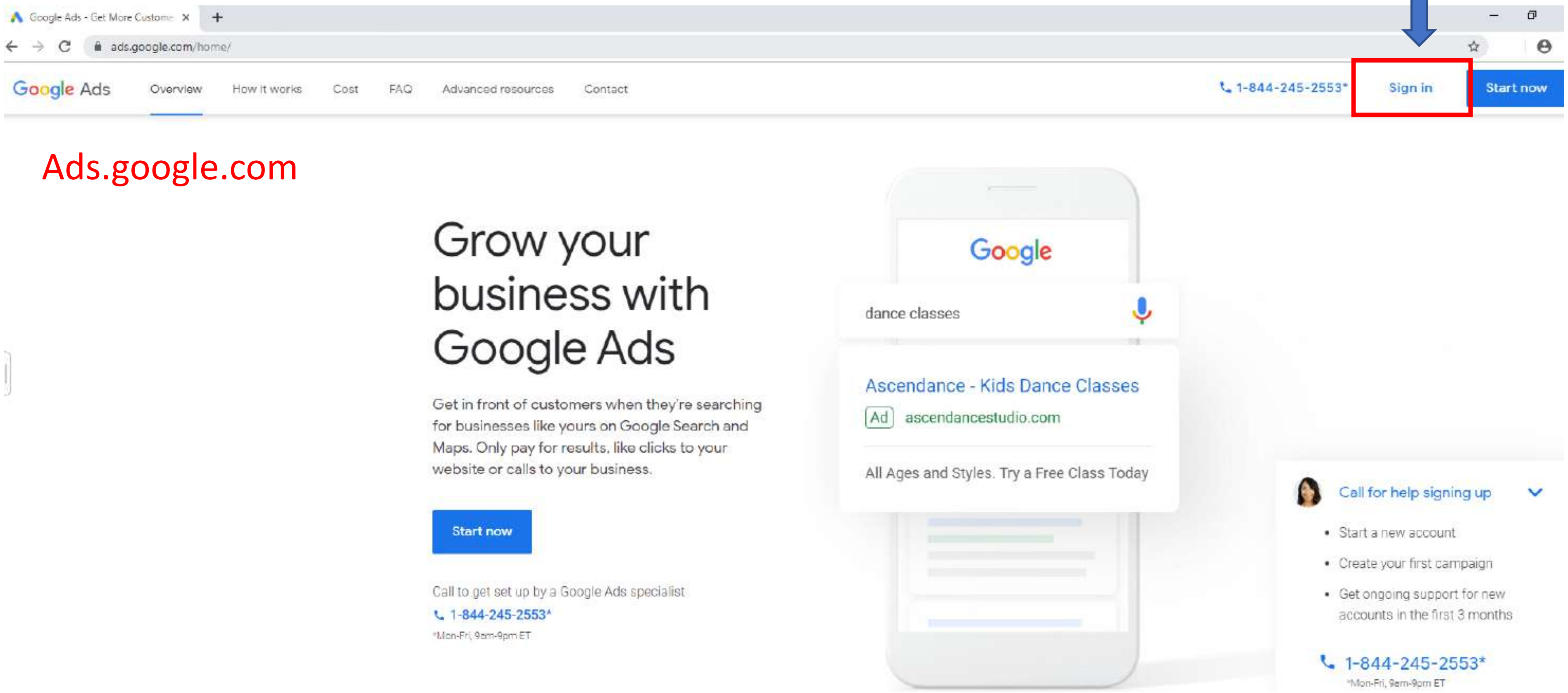
Francisco Villarroel Ordenes

October, 2022

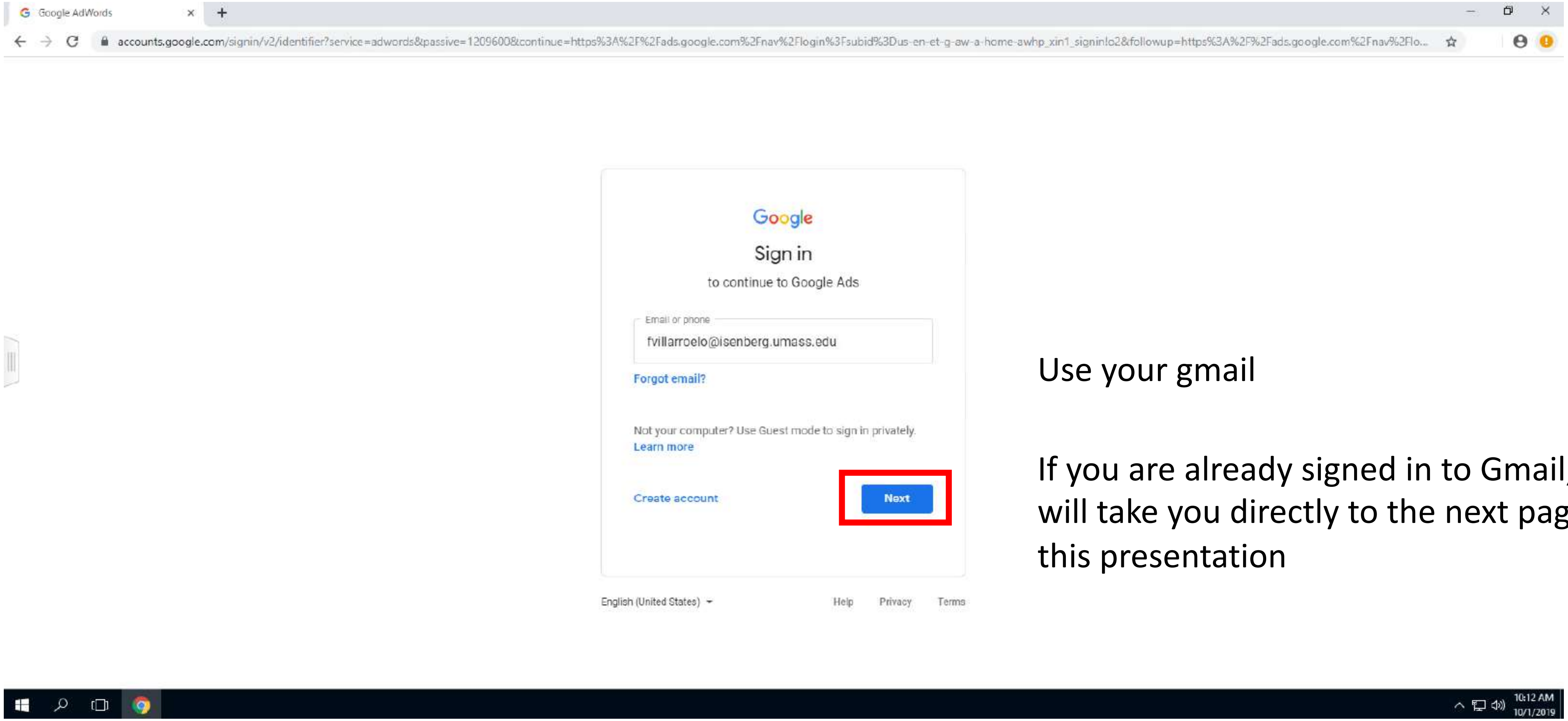
LUISS

Google Keywords

- In order to use google keyword you will need to sign in to ads.google.com (using your gmail account).
- Once signed in you must click "Experience with Google Ads?" and then "Create an Account Without a Campaign."
- If the webpage right after signing in does not offer the option "Experience with Google Ads?", then follow the process depicted in the next slides.

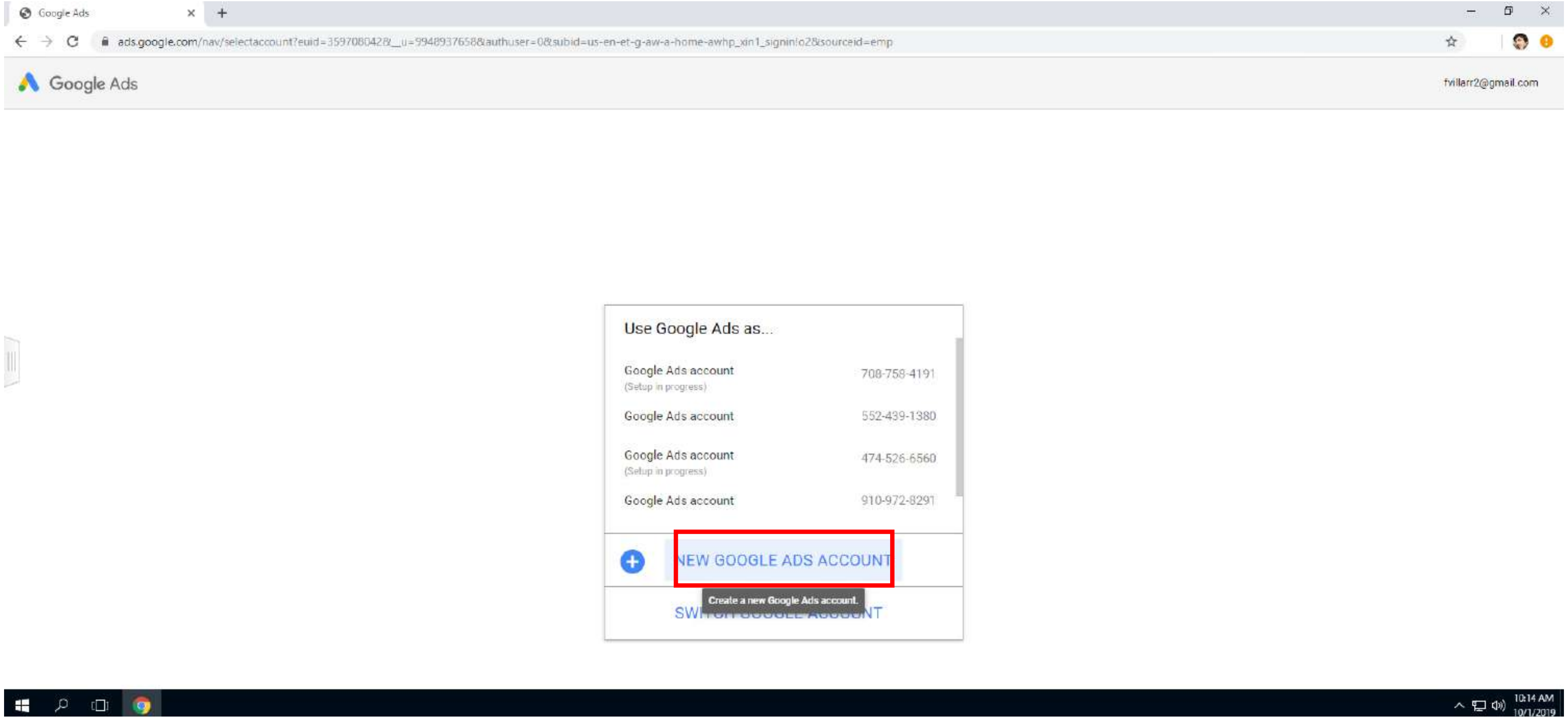


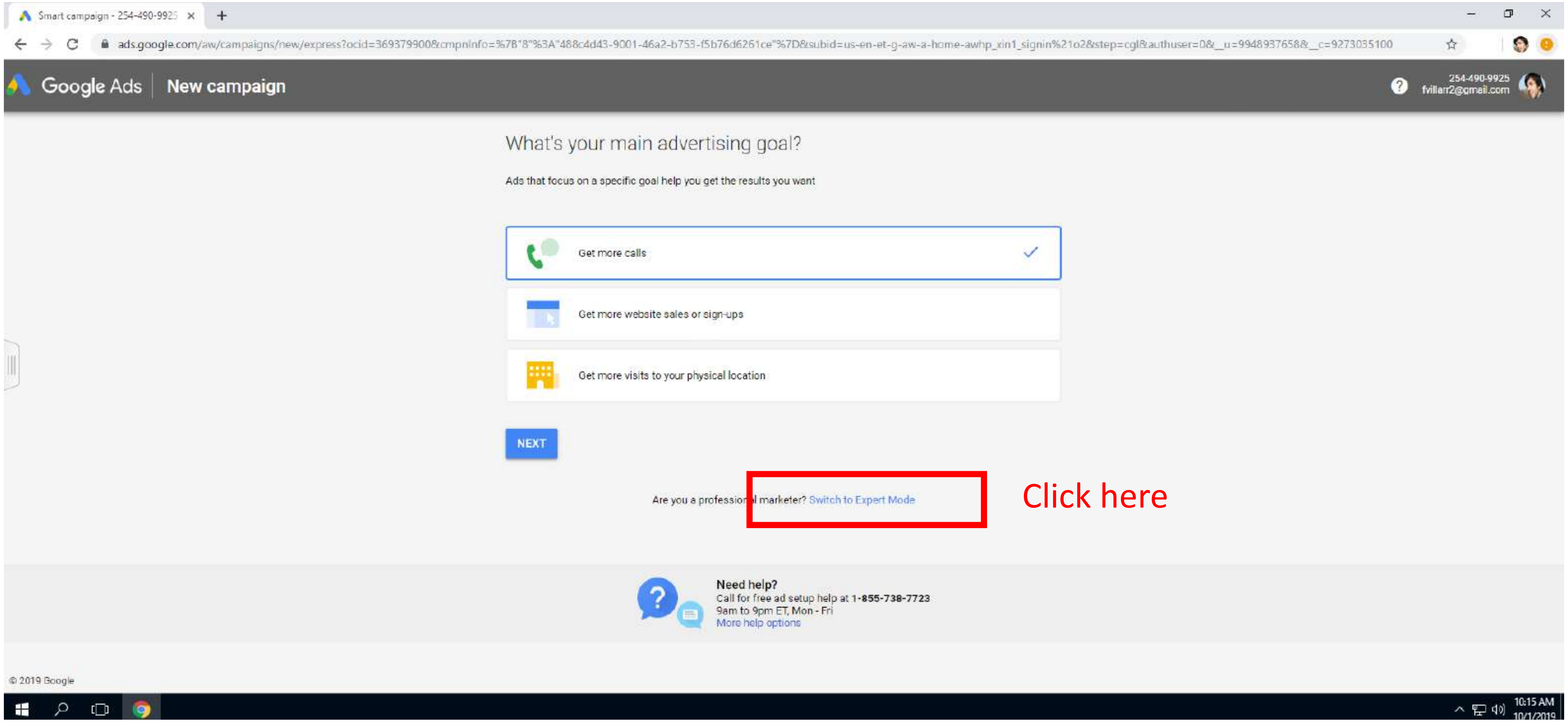
Ads.google.com



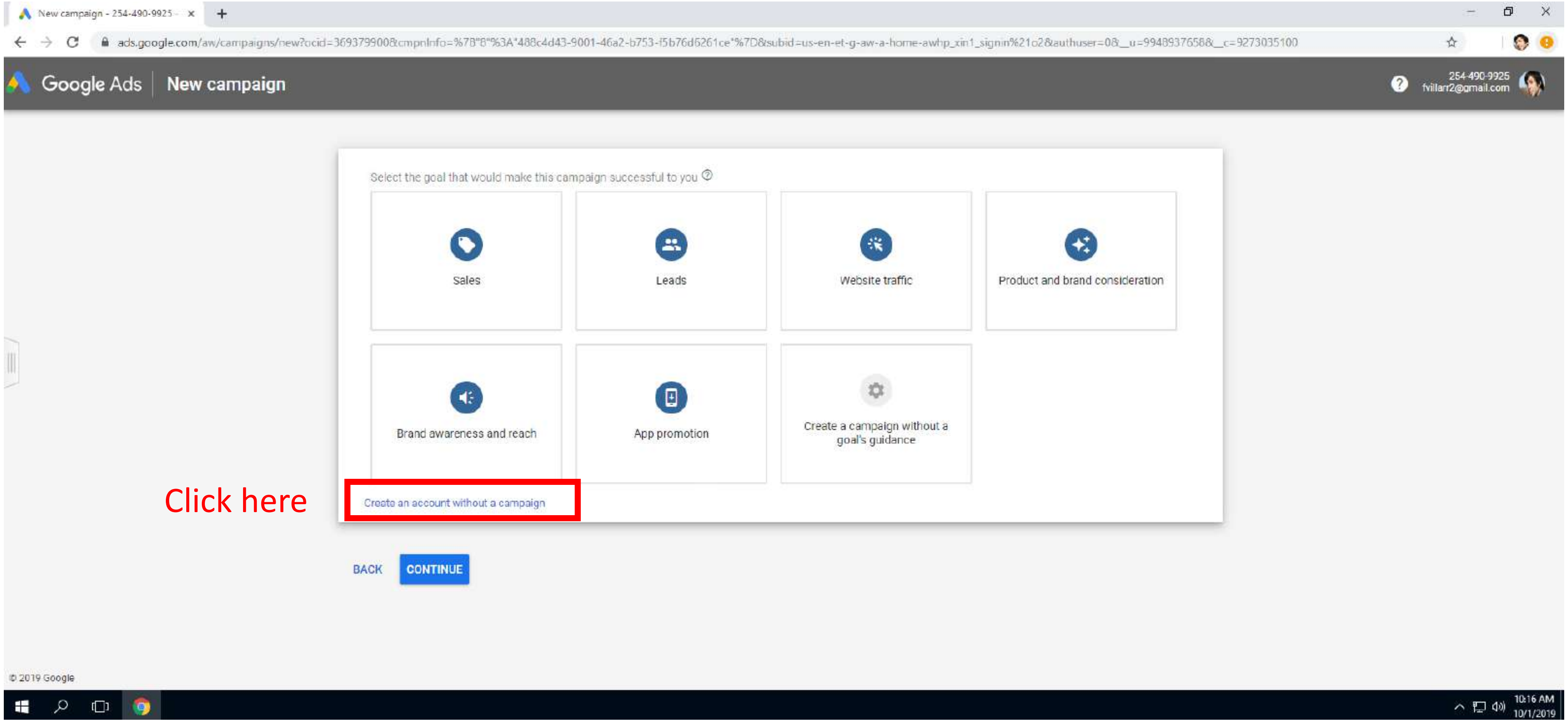
Use your gmail

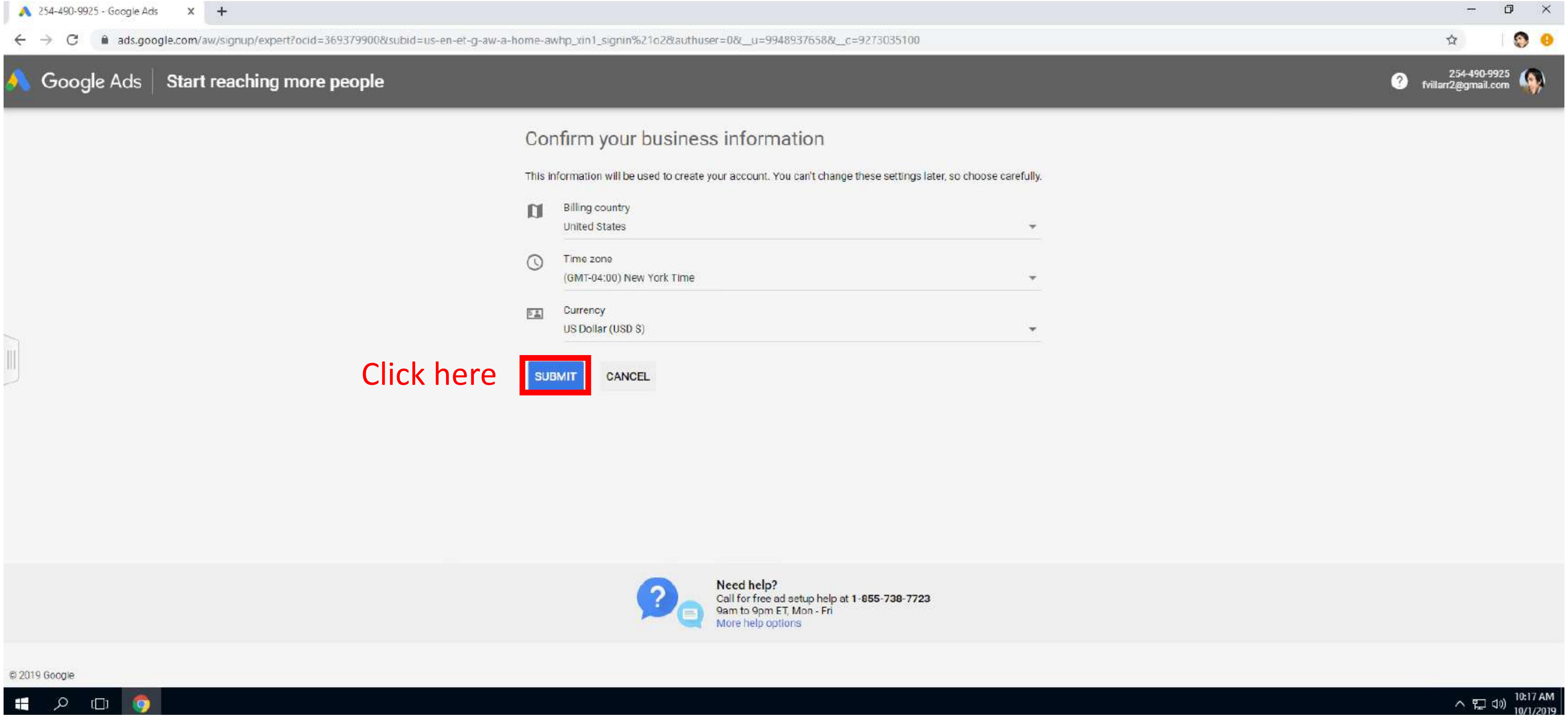
If you are already signed in to Gmail, it will take you directly to the next page in this presentation


Google Ads
Google Ads
fvillarr2@gmail.com
Use Google Ads as...
Google Ads account
(Setup in progress)
708-758-4191
Google Ads account
552-439-1380
Google Ads account
(Setup in progress)
474-526-6560
Google Ads account
910-972-8291
NEW GOOGLE ADS ACCOUNT
Create a new Google Ads account.

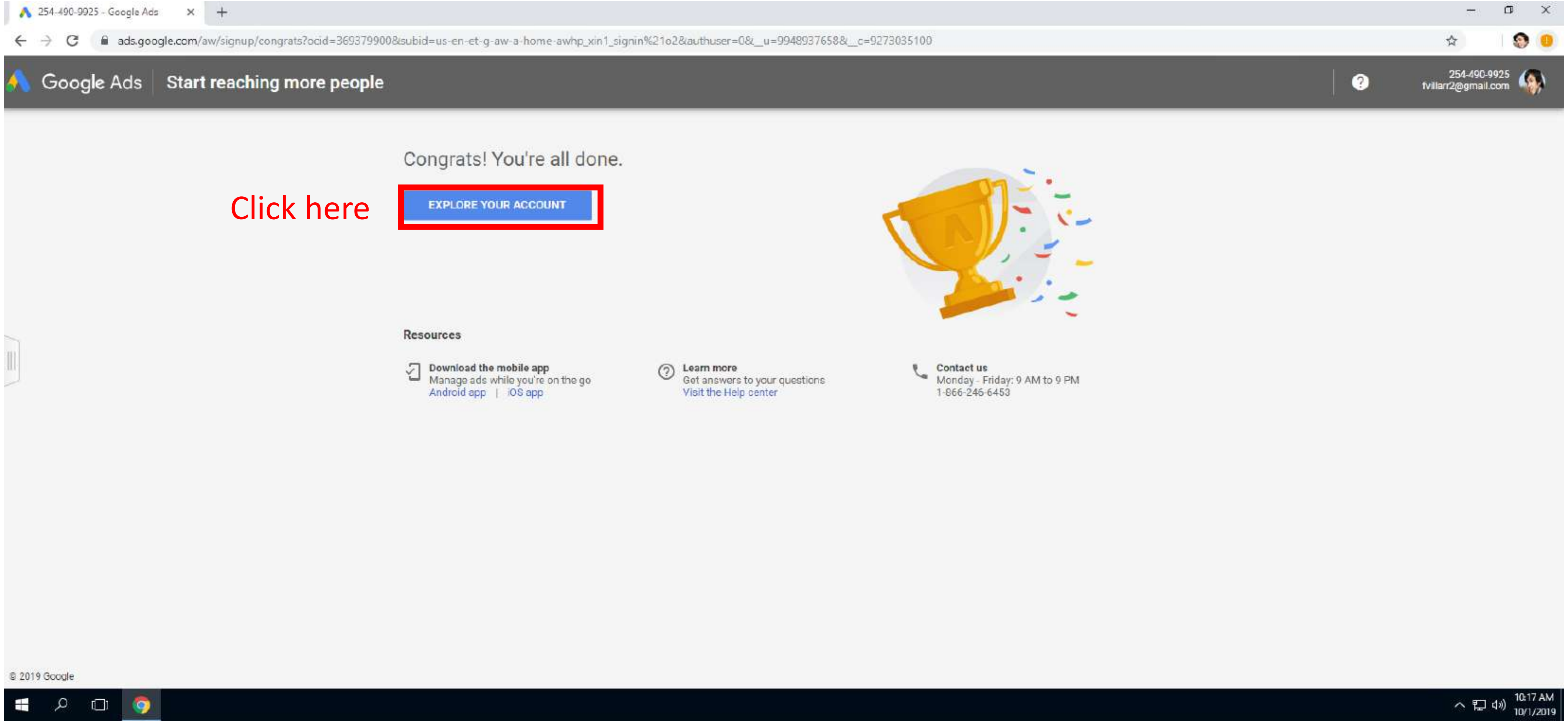

Smart campaign - 254-490-9925
Google Ads
New campaign
254-490-9925
What's your main advertising goal?
Ads that focus on a specific goal help you get the results you want
Get more calls
Get more website sales or sign-ups
Get more visits to your physical location
NEXT
Are you a professional marketer? Switch to Expert Mode
Need help?
Call for free ad setup help at 1-855-738-7723
9am to 9pm ET, Mon - Fri
More help options
© 2019 Google

Click here


New campaign - 254-490-9925
Google Ads
New campaign
254-490-9925
fvillarr2@gmail.com
Select the goal that would make this campaign successful to you
Sales
Leads
Website traffic
Product and brand consideration
Brand awareness and reach
App promotion
Create a campaign without a goal's guidance
Click here
Create an account without a campaign
BACK
CONTINUE
© 2019 Google
10:16 AM
10/1/2019

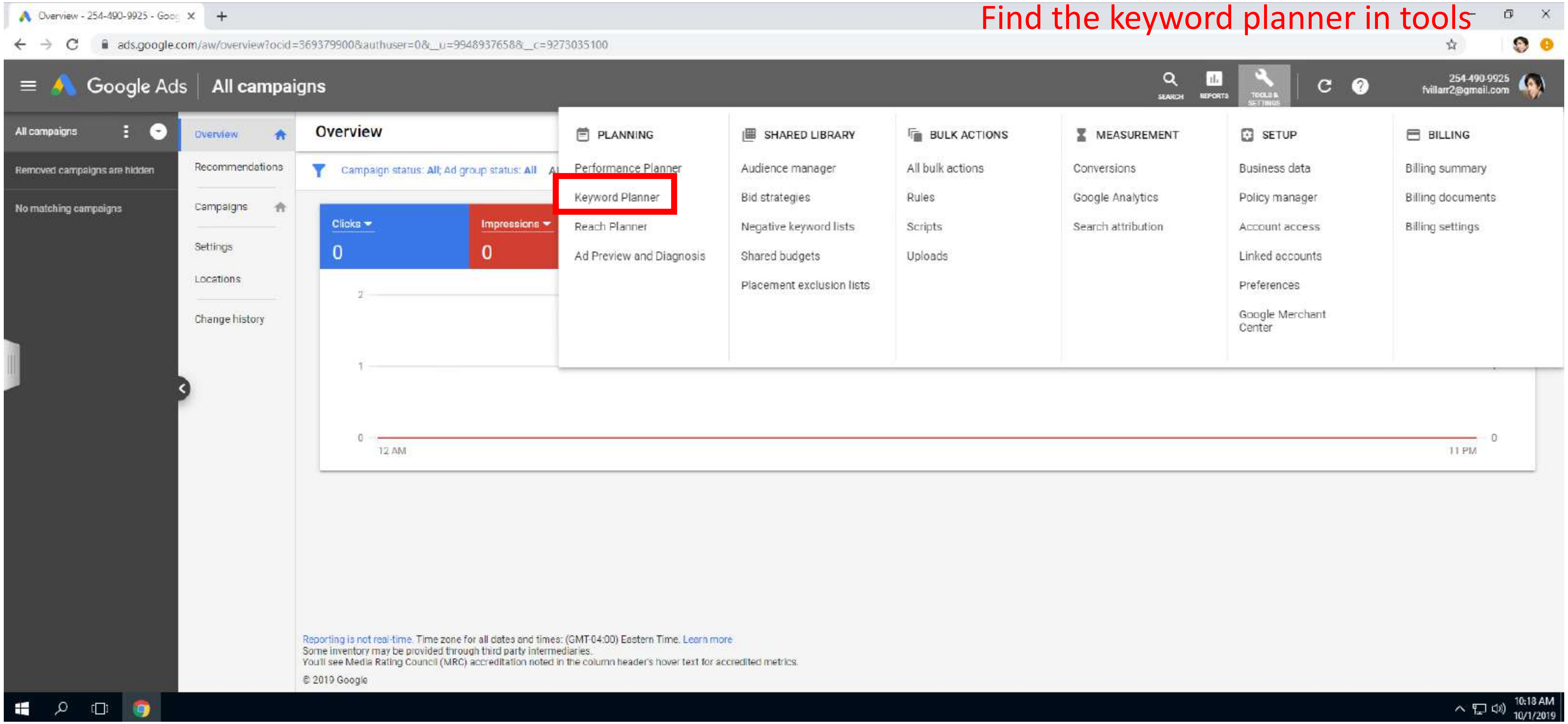


254-490-9925 - Google Ads
ads.google.com/aw/signup/expert?ocid=369379900&subid=us-en-et-g-aw-a-home-awhp_xin1_signin%21o2&authuser=0&__u=9948937658&__c=9273035100
Google Ads
Start reaching more people
254-490-9925
fvillarr2@gmail.com
Confirm your business information
This information will be used to create your account. You can't change these settings later, so choose carefully.
Billing country
United States
Time zone
(GMT-04:00) New York Time
Currency
US Dollar (USD $)
Click here
SUBMIT
CANCEL
Need help?
Call for free ad setup help at 1-855-738-7723
9am to 9pm ET, Mon - Fri
More help options
© 2019 Google
10:17 AM
10/1/2019

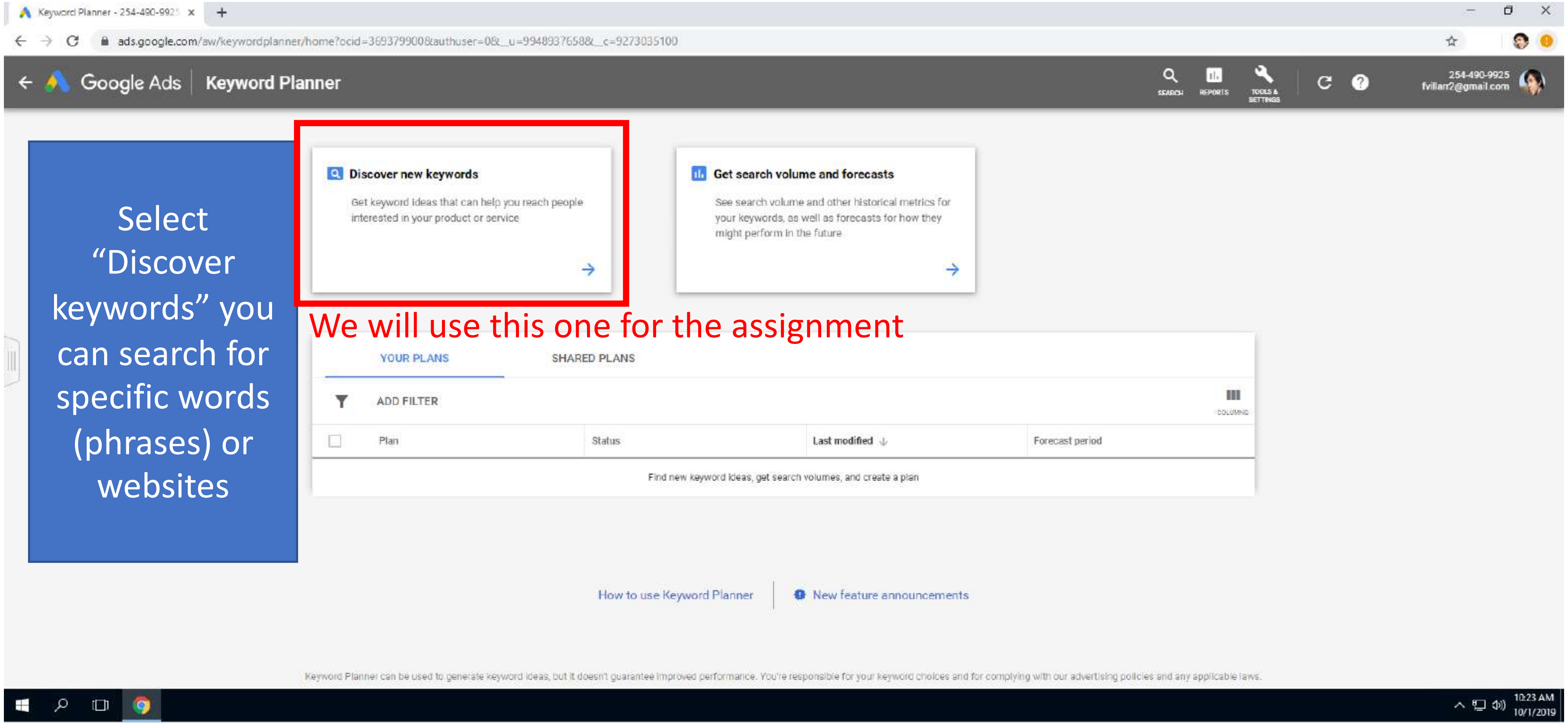


254-490-9925 - Google Ads
ads.google.com/aw/signup/congrats?ocid=369379900&subid=us-en-et-g-aw-a-home-awhp_xin1_signin%21o2&authuser=0&__u=9948937658&__c=9273035100
Google Ads
Start reaching more people
254-490-9925
fvillarr2@gmail.com
Congrats! You're all done.
Click here
EXPLORE YOUR ACCOUNT
Resources
Download the mobile app
Manage ads while you're on the go
Android app | iOS app
Learn more
Get answers to your questions
Visit the Help center
Contact us
Monday - Friday: 9 AM to 9 PM
1-866-246-6453
© 2019 Google
10:17 AM
10/1/2019

Find the keyword planner in tools



Start practicing keyword search



Thanks

Unit 4 (Week 5): Content Marketing & Explanatory Analytics (Part 2)

Course: Business and Marketing Analytics

Francisco Villarroel Ordenes

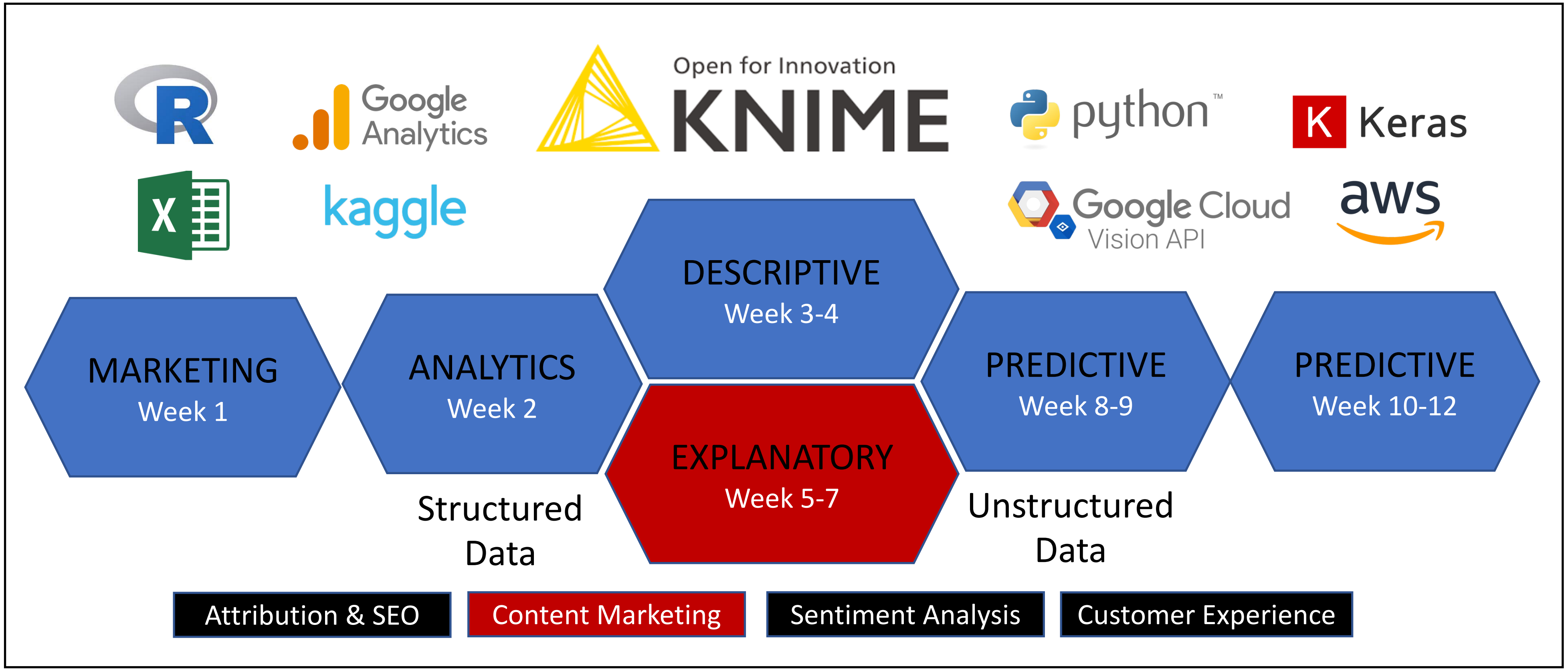
October, 2023

LUISS

Recap

Unit 4: Content Marketing and Explanatory Analytics

- Brand Equity
- Integrated Marketing Communication
- Content Marketing
- Social Media Marketing



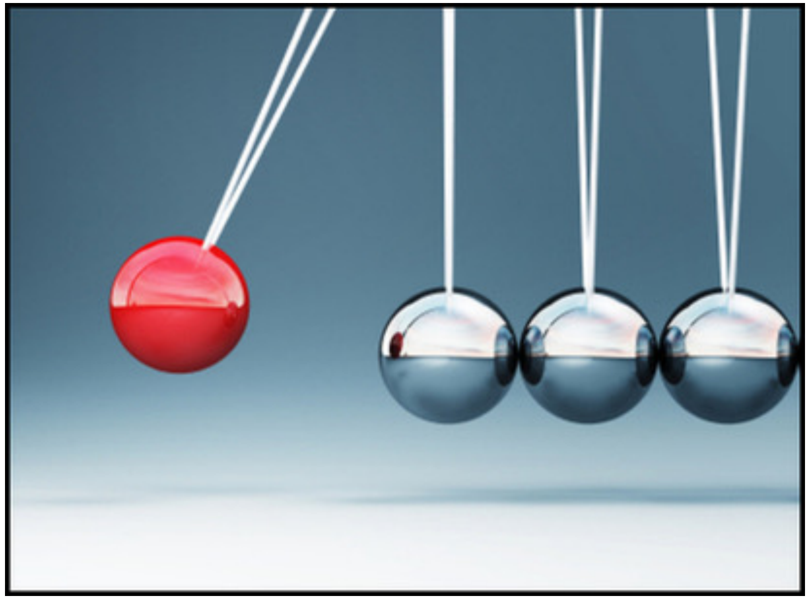
BUSINESS & MARKETING ANALYTICS FRAMEWORK

Objectives

- Applying explanatory analytics techniques in the field of social media content (this time not necessarily marketing).
- Introduction to conceptual models and hypothesis development (moderation & mediation)
- Introduction to regression models for count data (Negative Binomial)
- Results and Implications

Explanatory Analytics

- Diagnostic explanatory models that estimate relationships between variables and allow for hypothesis testing (Wedel and Kannan 2016)
- Other authors (Leeflang et al. 2020) use the term “descriptive models” to refer to explanatory analytics. The define descriptive models as models that describe demand and or supply relations on markets models that give answer to questions such as which marketing instruments affect sales or engagement



What causal relation(s) to test?

We should start from a relevant **managerial problem** at the consumer, firm and/or societal level.

How do we start testing relations?

When using **deductive reasoning** we start from the literature and identify a set of variables and relations that we are interested in testing

When using **inductive reasoning** we start from the data and identify patterns, constructs and variables. Then we might identify relations amongst them

Explanatory Analytics & Text Mining: News about Trump and their impact in online sharing

The Washington Post

The New York Times

USA TODAY

Formalized Models (Leeflang et. al 2020)

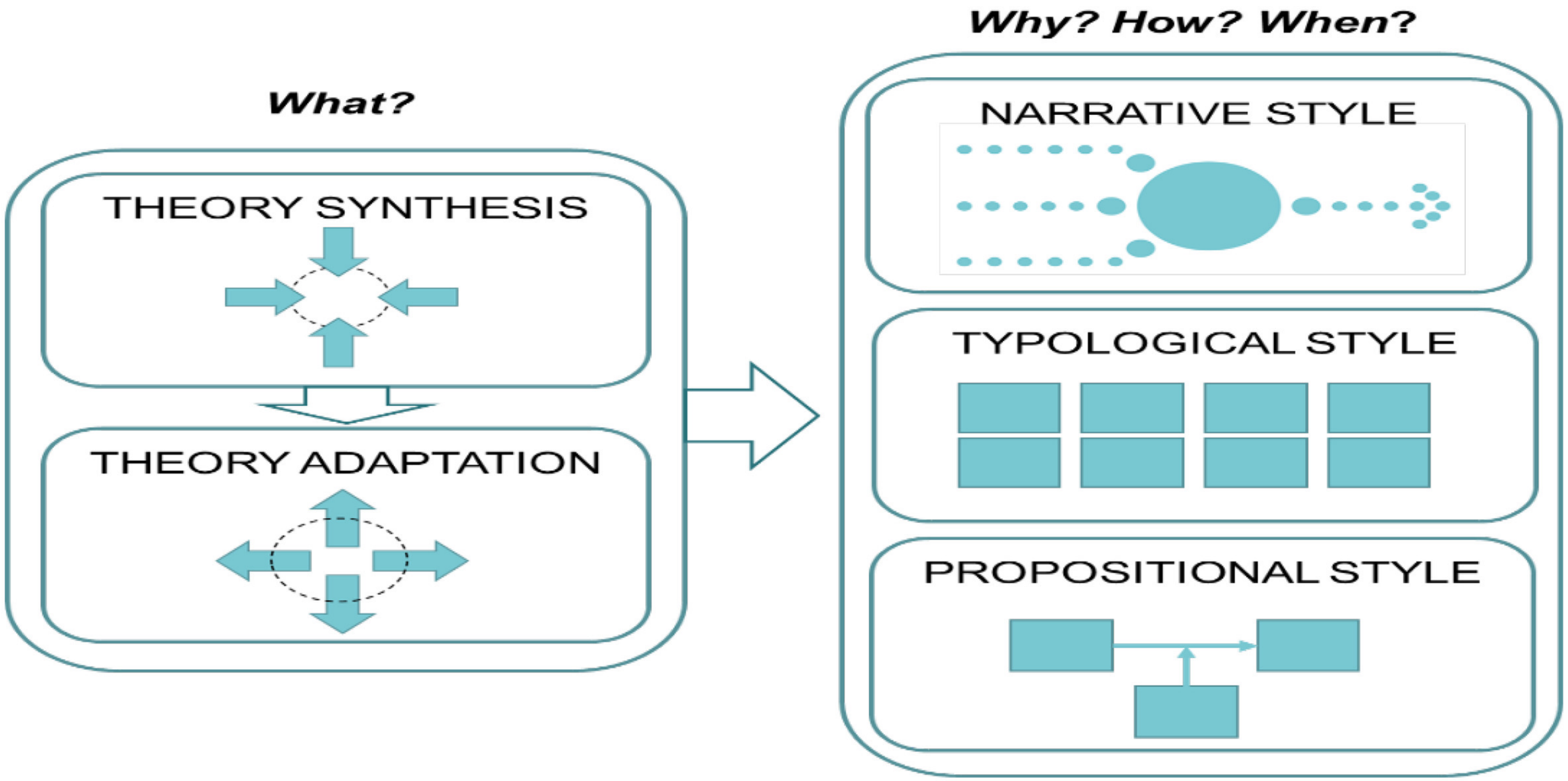
- Verbal Models: What a manager believes concerning marketing relations due to experience.
- Formalized Models
  - Diagram flows
  - Conceptual Frameworks
- Numerically Specified Models
  - Equation or set of equations that reflects marketing relations

Conceptual Frameworks and Theory

- The **way ideas are organized** to achieve a research project's purpose (Shields and Rangarajan 2013)
- **Informed by theory**, conceptual frameworks allow researchers to identify a sound set of relationships between managerial constructs of interest (customer satisfactions -> sales) (Lindgreen et al. 2020)
- A theory is a statement of **concepts and their interrelationships** that shows **when, how and/or why** a phenomenon occurs”
- The development of conceptual frameworks is **non-linear & iterative**

Types of Conceptual Framework

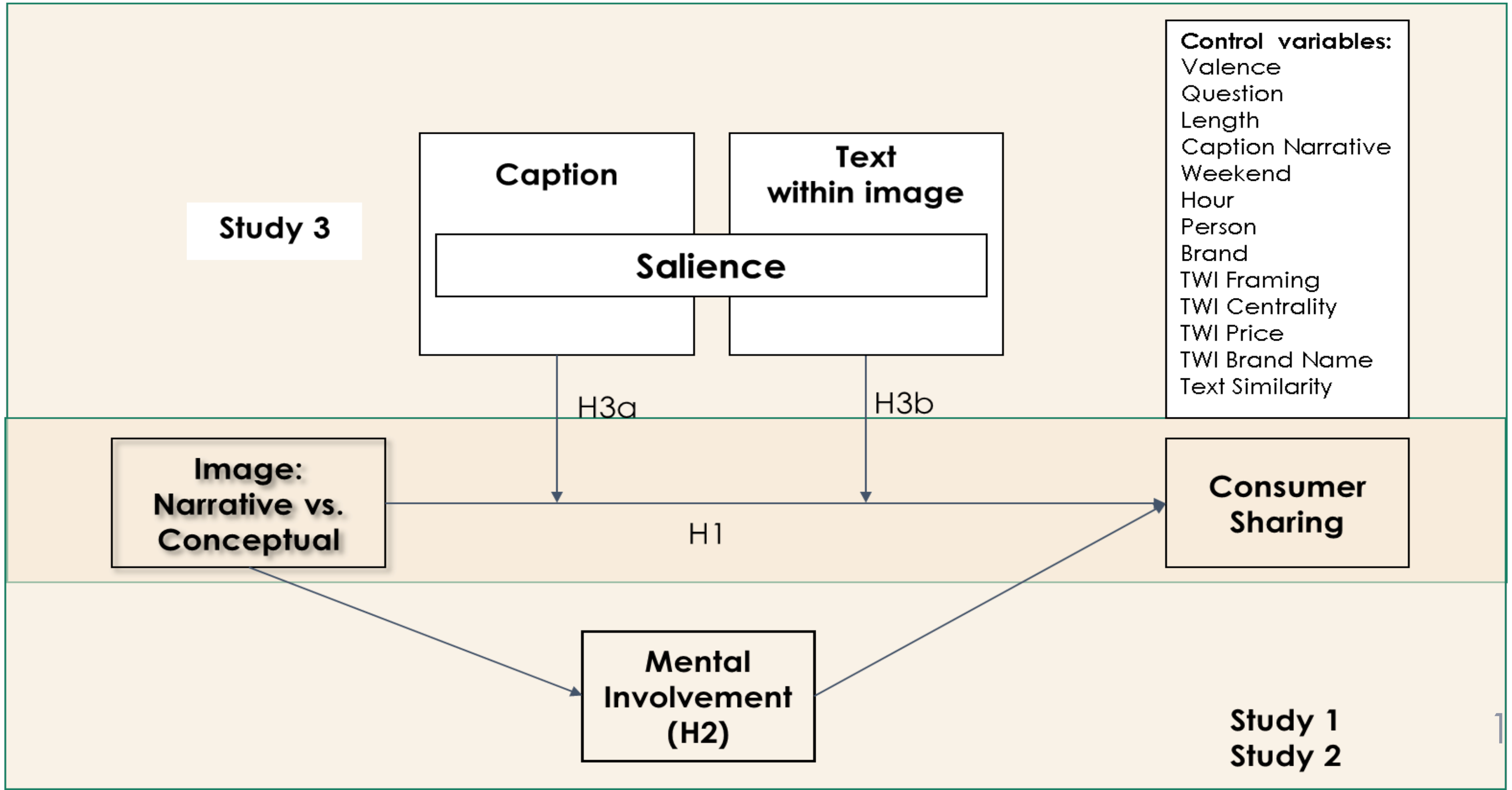
All types of conceptual framework aim answering why, how and when questions (Lindgreen et al. 2020)



Types of Conceptual Framework (1)

Proposition Based Style: The statement of theoretical propositions that introduces new constructs and cause-effect relationships

Farace et al. (2020): "The Text + Image Formula in Social Media Brand Posts that Really Works"

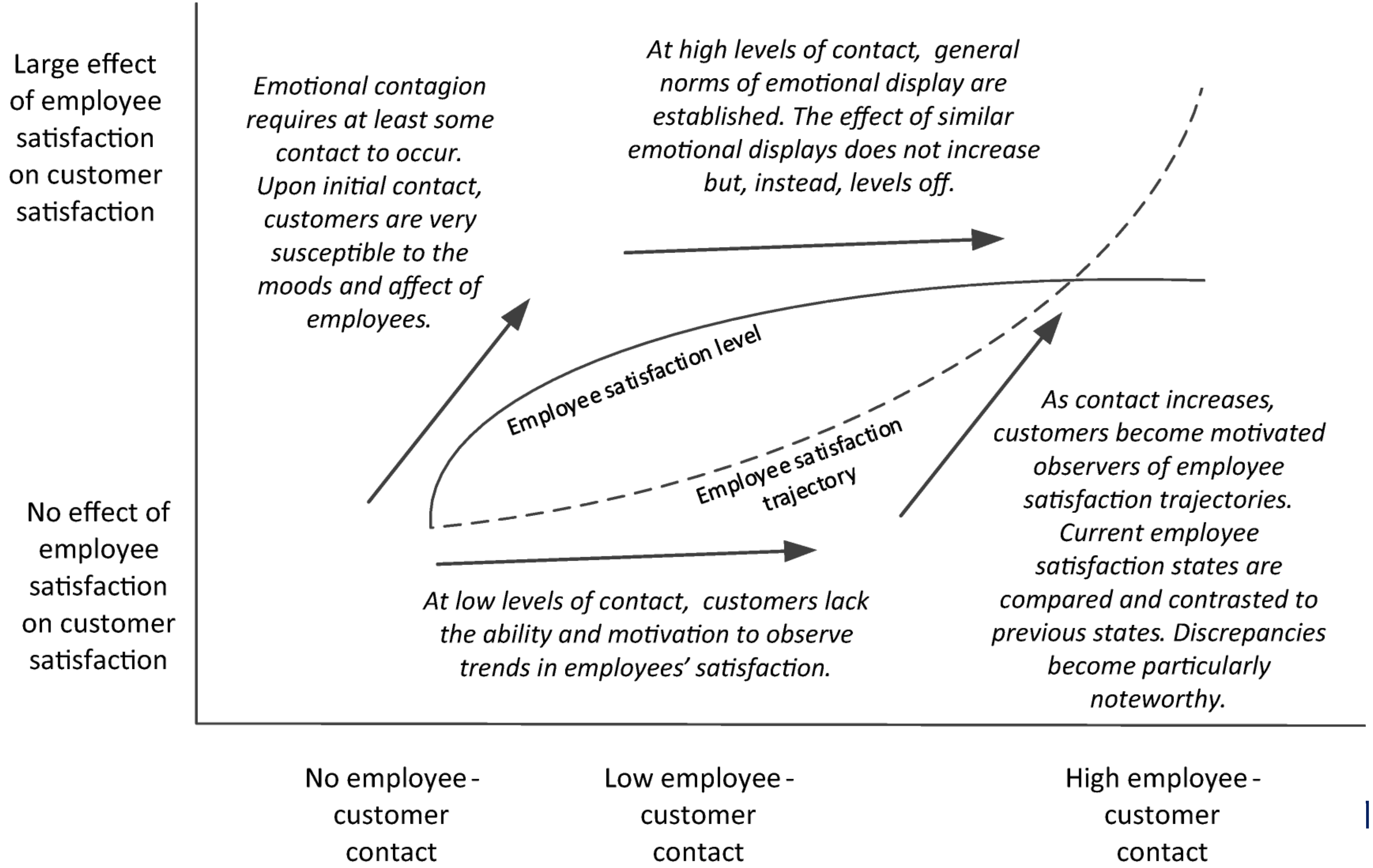


LUISS

Types of Conceptual Framework (2)

Narrative Based Style: The specification of a process model that lays out a set of mechanisms explaining events and outcomes.

Wolter et al. (2019): "Employee satisfaction trajectories & their effect on customer satisfaction & repatronage intentions"



Types of Conceptual Framework (3)

Typology Based Style: The specification of a typology that interrelates different dimensions to flesh out new constructs and causal interactions.

Table 1. Evolution of Consumption: A Psychological Ownership Framework.

Dimension of Change	Threats to Psychological Ownership	Transfers of Psychological Ownership	Opportunities to Preserve Psychological Ownership
Legal ownership to legal access. Personal ownership of private goods is replaced with temporary access rights to use collectively consumed goods and services.	*Fractional ownership.* Bundle of rights associated with good divided among agents holding property rights to use, profit, change, or transfer ownership. ❖ Emphasize liquidity and economic value. *Impermanence.* Consumers no longer expect to keep goods—they assume goods will be returned, impairing reference-point shift to owner ("My…"). ❖ Extend/guarantee duration and consistency of consumption experience.	*Collective consumption.* Ownership felt for private goods transfers to goods used collectively ("MINE" to "OURS"). Reduced importance of individual goods, potential contaminated by dissociative group associations. Psychological ownership transfers to consumer communities. ❖ Develop object history/intimate knowledge, encourage self-investment, deploy counterconditioning, and develop consumer communities.	*More consumer choice.* Improved preference-matching due to more (often immediately) available options, increases perceived control. ❖ Provide larger assortments, increase mass customization. *New channels for self-expression.* Social media and reputation systems integral to access-based consumption platforms provide new outlets for social signaling. ❖ Develop social media applications and marketing strategy, encourage microblogging, offer access to aspirational brands/goods with positive signal value.

Types of Conceptual Framework (3)

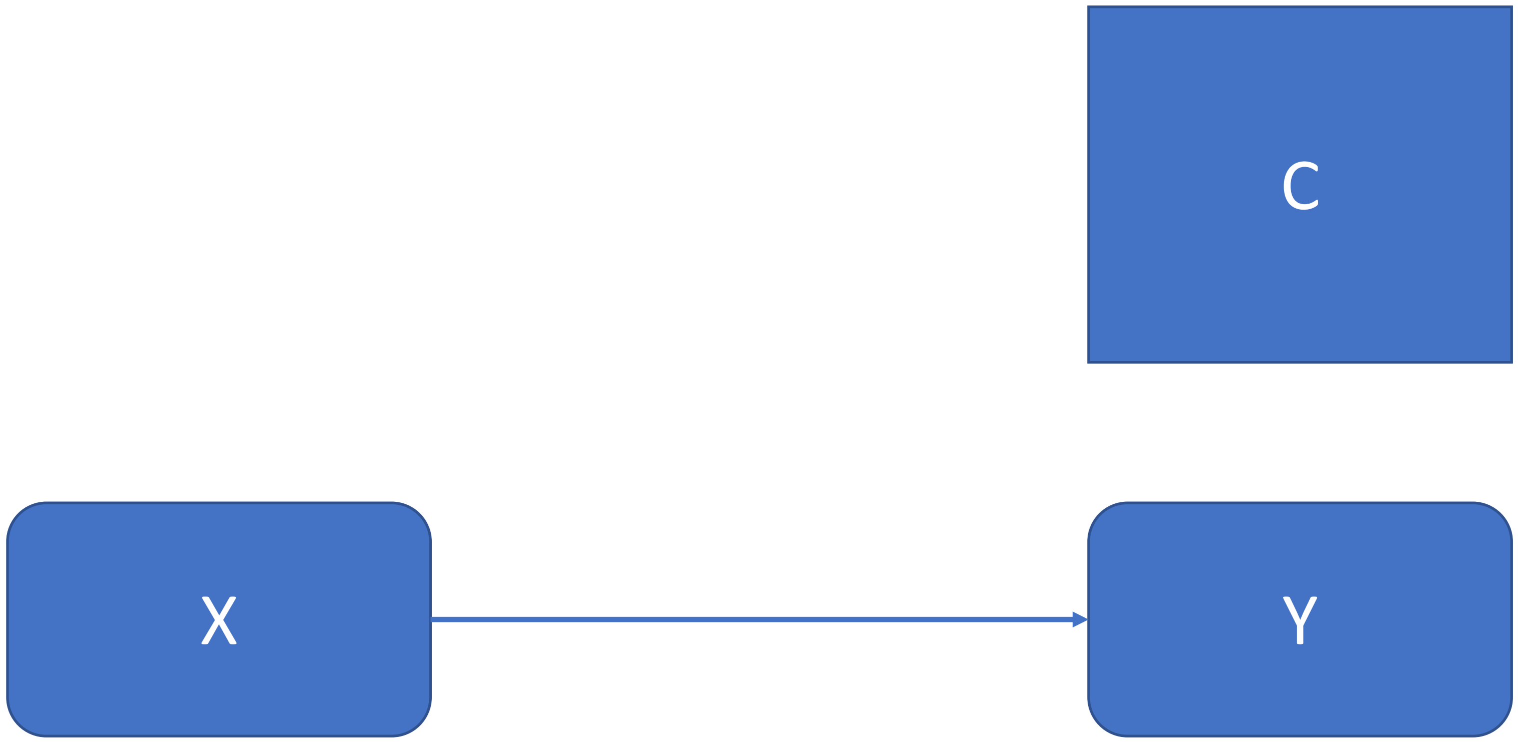
Table 1 (continuation; bottom part)

Material to experiential. Material goods are replaced with physical or digital experiential goods.	*Intangibility.* Consumers are less able to touch, hold, and physically manipulate experiential goods than physical goods. ❖ Develop haptic interfaces; interactive content; offer control over rate and timing of consumption; emphasize sensory features. *Reduced evaluability.* Ownership status is harder to determine (e.g., ownership of a vacation less clear than ownership of a vacation home). ❖ Make goods indexical connections–cues for personally meaningful events (e.g., cross sell physical goods, usage history reminders); gamification.	*Higher categorization level.* Category for which psychological ownership is experienced rises from individual goods to intermediary devices, platforms, and brands. ❖ Vertical integration, brand alliances, servitization, relationship marketing, intermediary device personalization.	*Greater self-identification.* Experiences are easier to integrate with self-concept than material goods (e.g., experiential purchases may generate more positive self-signals). ❖ Leverage identity marketing (e.g., "I am a skier" > "I own skis").

Notes: ❖ = recommended marketing actions to manage psychological ownership threats, transfers, and opportunities.

We will focus on proposition-based frameworks (ideal for explanatory analytics)

Main Effect (Proposition Based Style)



- Independent Variable (X)
- Dependent Variable (Y)
- Control Variables (C)

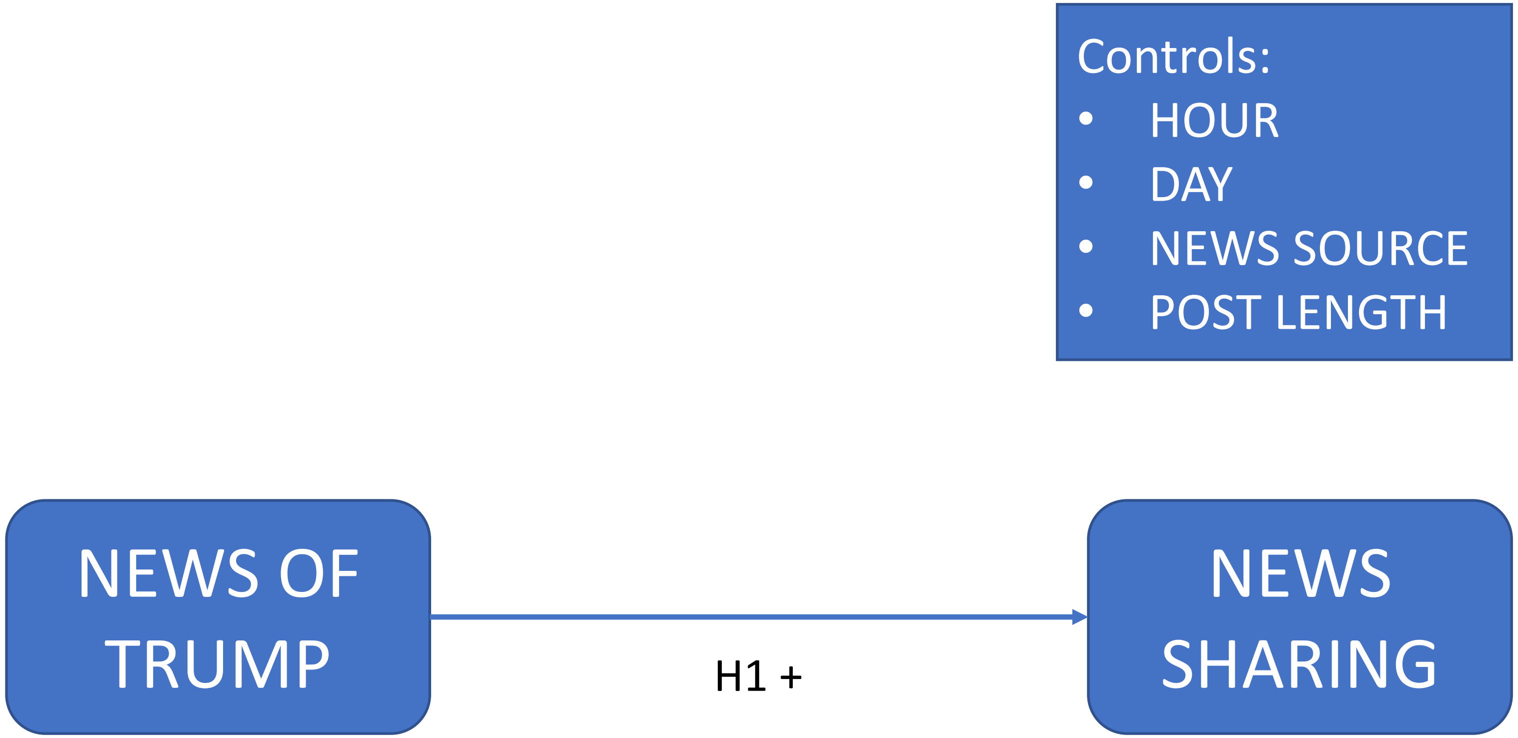
Main effect

- Imagine we want to test a hypothesis concerning news media containing information about Donald Trump and its effect on retweeting. How would you hypothesize?

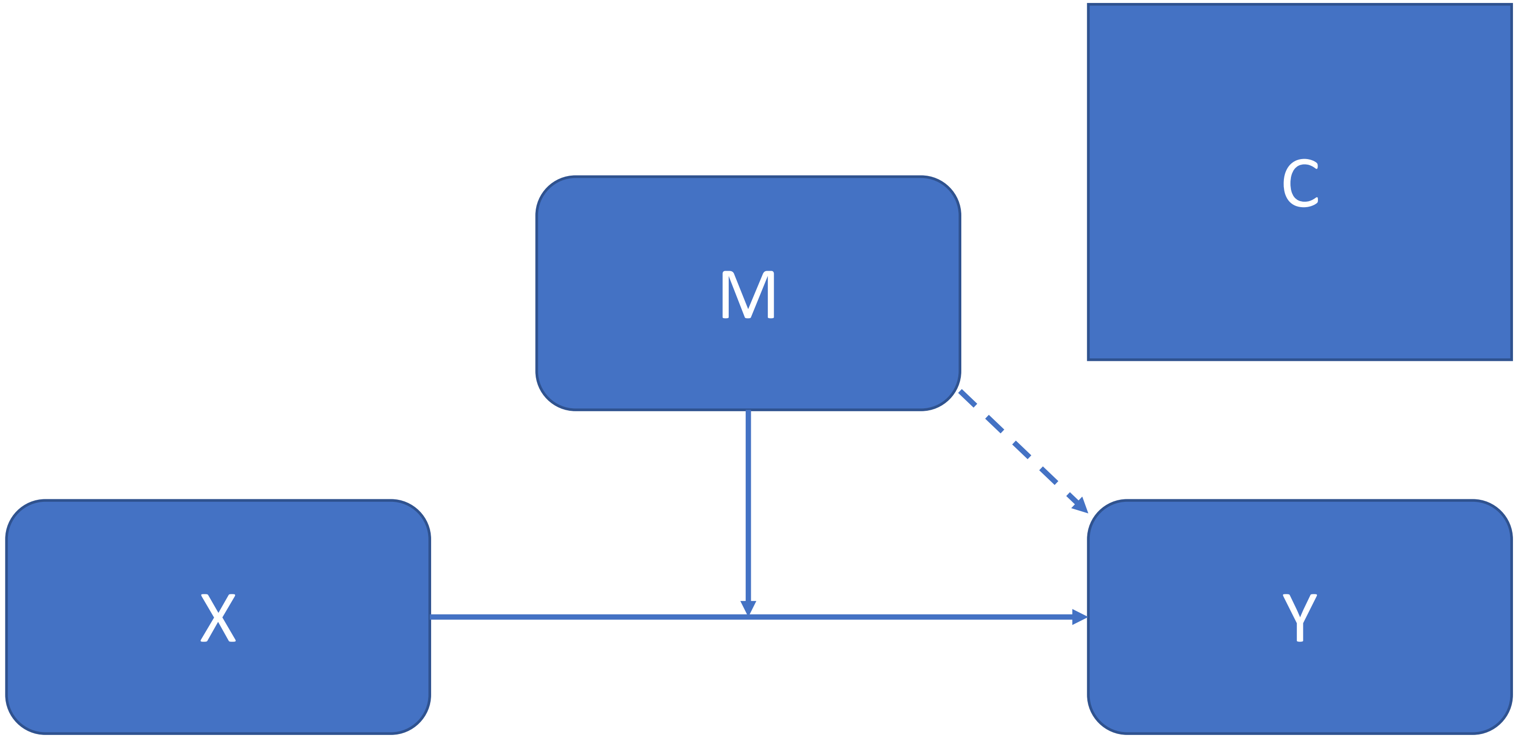
- We expect that news containing information about Donald Trump will be more relevant nowadays because of the US Election. In addition, because of his high presence on Twitter, news about Donald Trump should attract attention (Build up without theory)

H1 = News media tweets containing information about Donald Trump will be more shared than tweets non-related with Donald Trump

Main Effect


Controls:
• HOUR
• DAY
• NEWS SOURCE
• POST LENGTH
NEWS OF TRUMP
H1 +
NEWS SHARING

Moderation Effect



- Independent Variable (X)
- Dependent Variable (Y)
- Control Variables (C)
- Moderator Variable (M)

Moderator Effect

- A moderator is a variable that specifies conditions under which a given IV is related to a DV.
- Moderation implies an interaction effect, where introducing a moderating variable changes the direction or magnitude of the relationship between two variables.
- A moderation effect could be:
  - **Enhancing**, where increasing the moderator would increase or strengthen the effect of the IV on the DV
  - **Buffering**, where increasing the moderator would decrease or weaken the effect of the IV on the DV

Moderation effect

- What could be a moderator in our current model? Things that attenuate or strengthen sharing of news about Donald Trump
- **Negative Tone:** Drawing on "Negative Bias Theory" (things of more negative nature have more impact on people's behavior & cognition), we believe that greater negative news content will be more shared
- **TRUMP & Negative Tone:** The negative associations of TRUMP news, together with more negative news wording, will result in a double negative content. Excess of negative information will overload people with negativity, resulting in lesser sharing

H2 = Greater negative tone in news media tweets will result in greater consumer sharing

H3 = Greater negative tone in news media tweets will **weaken** the positive effect of TRUMP news on consumer sharing

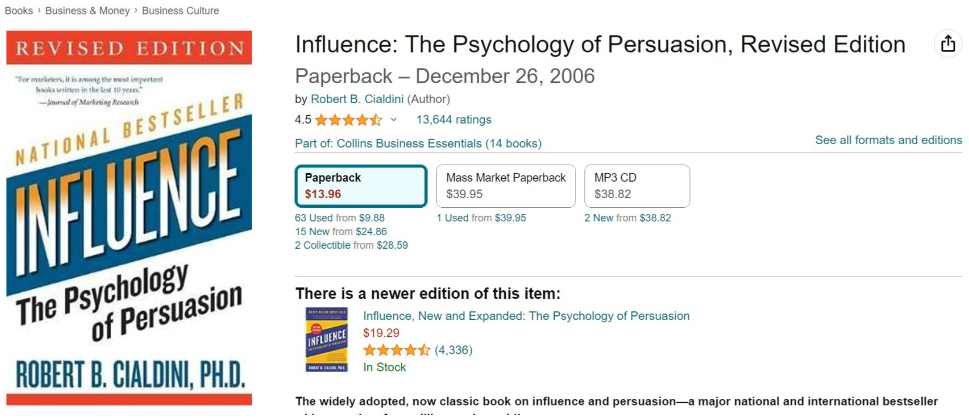
A Preview? “Tag cloud of Trump News”

Mid-term evaluation (N=59)

- **POSITIVES**
- Diversity and complementarity of things. Course Organization. Theory, Labs, Assignments
- Guest Speakers!
- Regular quizzes to spread out the work, not too stressful
- Helpfulness and friendliness of the teaching team
- Interactivity and practical value of the course
- Respect for the syllabus

Mid-term evaluation – Negatives

- “Papers are not necessarily suited to marketing”
- Multiple choice not fair to evaluate
- More “scientific approach” and less abstract
- Students that use R or Python (low code not used in industry; specially those in AI and ML roles)

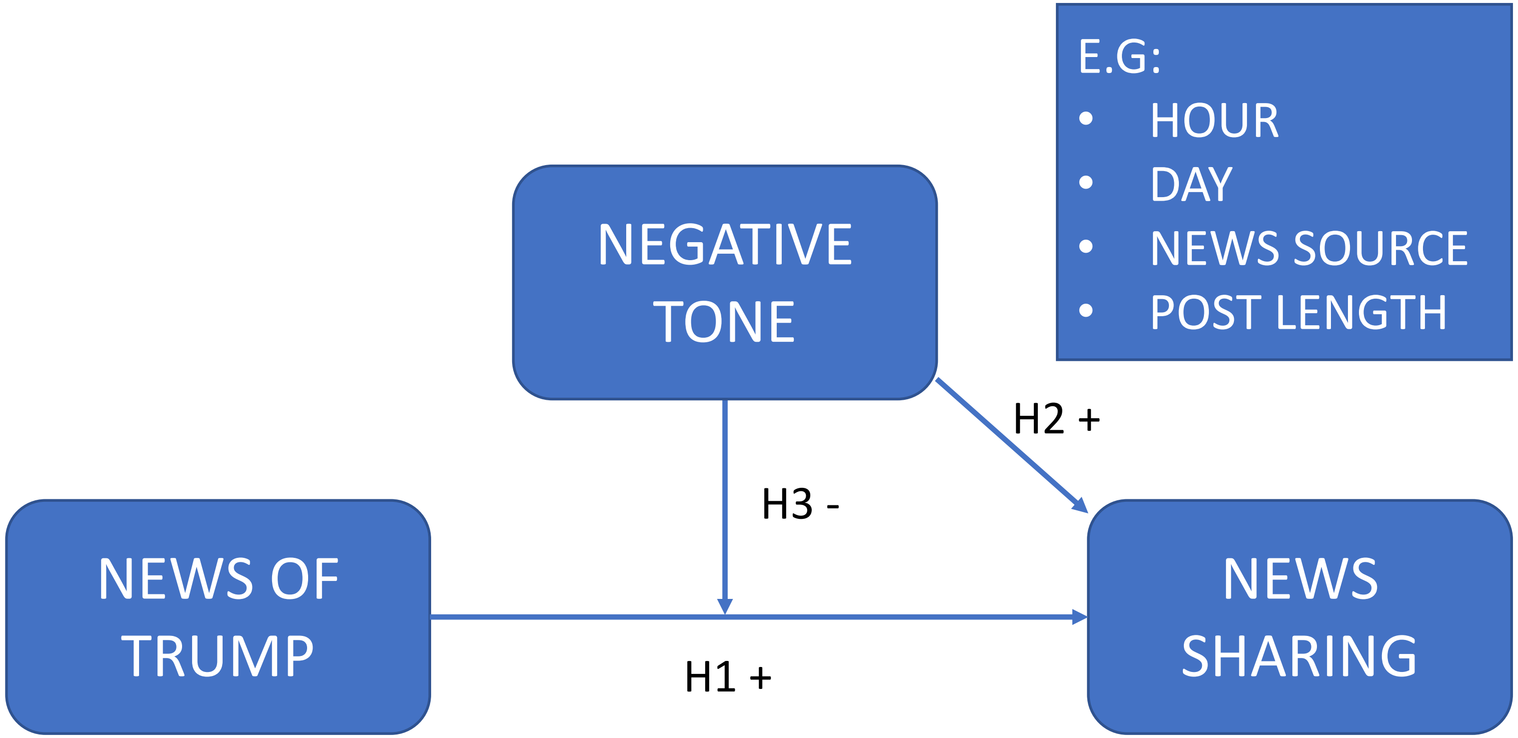


- 15% participation grade unfair for “shy” students
- More time to work on KNIME because people doesn’t know about programing
- Little time for questions in guest sessions
- Course structure with weeks and slides

Midterm evaluation – reminders and changes

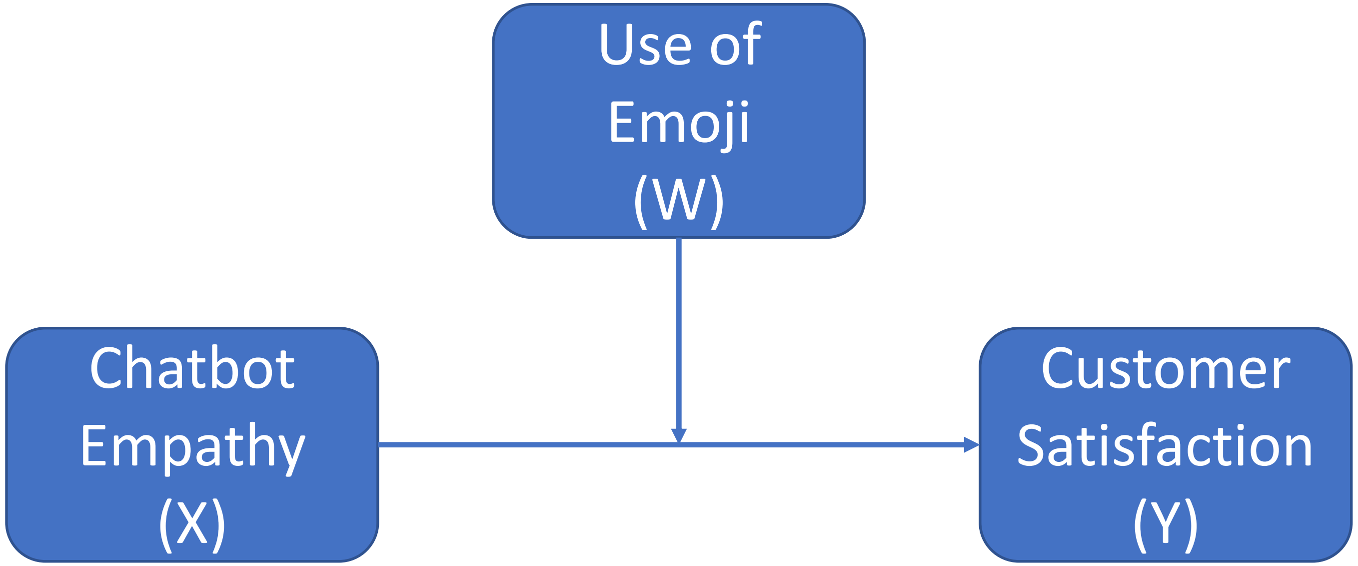
- Participation (15%). Just attending 26. It increases not only by asking questions, but by submitting insights (what you learned). Being “shy” might be something to work on
- Articles. We will indicate page numbers on where to focus
- Individual and group evaluation (50% - 50% is a good approximation of real life)
- KNIME labs: Remember we also offer student hours (Monday, 16:00-19:00, presence or digital is possible)
- In the last assignments teams are allowed to us R or Python to solve it if they want/can

Moderation Effect



Moderation Effect; Other Examples

- When (i.e., under what condition) is the relationship between chatbot empathy and customer satisfaction positive?
- For example, is chatbot use of emojis increasing the positive effect of chatbot empathy on perceived customer satisfaction?



- Independent Variable (X)
- Dependent Variable (Y)
- Moderator Variable (W)

A basic distinction (Leeflang et al. 2015)

Two types of Models in Marketing:

1. Models made for the primary purpose to support decision making of a specific marketing manager
    - Primary target audience are managers
    - It leads to case specific insights
    - Data might be focused on an industry or brand
2. Models that aim to advance general marketing knowledge
    - Primary target audience are marketing scientists/researchers
    - It leads to generalizable insights in marketing phenomena
    - Data should involve several brand/industries and periods

The Model Building Process (Leeflang et al. 2015)

1. Opportunity Identification: Managerial Problem.
2. Model Purpose: What problem should be solved?
3. Model Scope: What are the boundaries of it.
4. Data Availability: In this case Tweets!
5. Model Specification: Equation.
6. Estimation: OLS, ML, and others.
7. Validation: Make sure that it works (model fit, accuracy, etc.)
8. Cost-benefit consideration: Is it worth investing on it?
9. Use: How it can be used by a manager for decisions?
10. Updating: How to keep it working over time.

Model Criteria (Leeflang et al. 2015; Chapter 2)

- *Simple:* keeping the number of variables small, and only keep the relevant phenomena in the model
- *Evolutionary:* starting simple and expanding as time goes on (contributes to managers step-by-step understanding)
- *Complete:* account for all important variables. This will be relative to the problem, user and organization
- *Adaptive*: need to be updated regularly due changes in environment (e.g., parameters and structure when a new incumbent in the market, or changes in consumer base)
- *Robust*: a quality characteristic which makes it difficult for a user to obtain bad answers. The model builder needs to identify relevant constructs, define valid measures, specify meaningful functional forms (e.g., non-linearities), and accommodate appropriate interaction effects

How to test conceptual frameworks with Explanatory Analytics/Models

Several **regression models** can be used to test a conceptual model. The choice will depend on the assumptions of the model and generally the characteristics of the dependent variable.

- Linear (DV: Scale or Ratio)
- Logistic (DV: Yes or No; 1 or 0)
- Ordinal Logit (DV: Star Rating)
- Multinomial (DV: Categorical)
- Poisson & Negative Binomial Regression (DV: Count)
- Ridge/Lasso Regression (for multicollinearity)
- Bayesian Regression (probability based)

Linear Regression Models to Assess Causality

Correlations only provide insight into bi-variate relations, what we are looking for, is causal or predictive relations between independent and dependent variables in a model of the real world that allows us to control for as much noise as we can. For this we need to conduct a regression model.

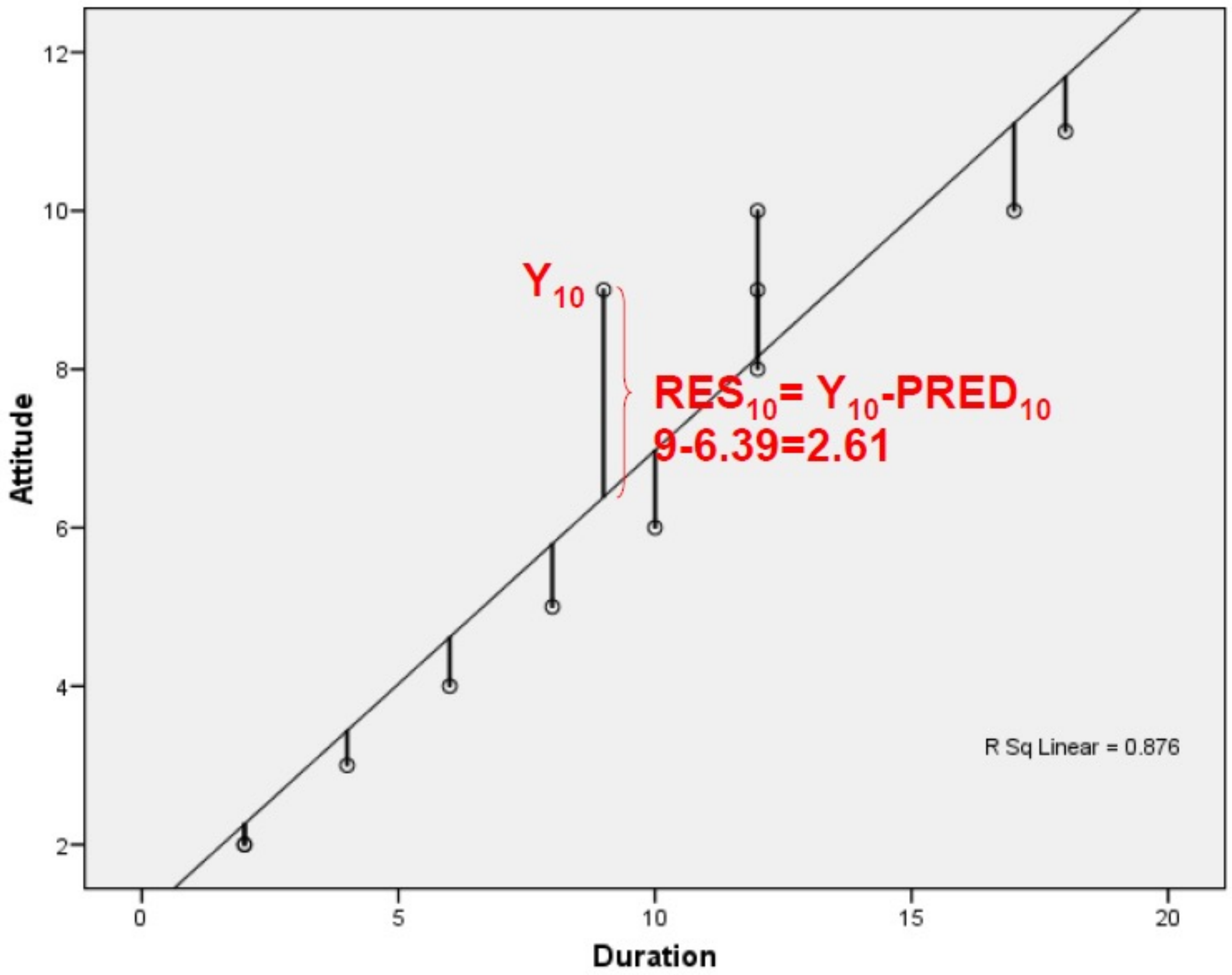
Model: $Y_i = \alpha + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \mu_i$

Estimation: *Ordinary Least Squares* (OLS; tailored for linear models) or *Maximum Likelihood* (ML; tailored for non-linear models). OLS aims to minimize the distance between Y and the predicted Y', ML is a likelihood maximization.

Fit Measures (how good your model is):

- *Adjusted* R^2 for OLS linear regression. Ranges from 0 (bad fit) to 1 (good fit
- *Log Likelihood, AIC and Pseudo R Square* for non linear models and ML

How it works (linear models)?



Assumptions of Multiple Linear Regression (Wooldridge 2018; Chapter 3)

Under these assumptions parameters are unbiased:

1. **Linear in parameters.** There is a linear relationship between independent variables and the dependent variable. Yet, this is flexible because we can use different functions for dependent variable (e.g., log).
2. **Random sampling.** We have a random sample of n observations, and All Yi values are independent
3. **No perfect collinearity.** There does not exist any exact linear relationship between the Xi's (assumption of no multicollinearity). This can be assessed (preliminarily with a correlation table, or with VIF indicator)
4. **Zero conditional mean.** The ERROR "μ" has am expected value of zero given any values of the independent variables. It can fail if the relationship between explained and explanatory variables is misspecified (e.g., by forgetting to include a quadratic term). Omitting a variable that is correlated with any of the predictors would violate this assumption (**Endogeneity).**

***Homoskedasticity:** the error μ has the same variance given any values of the explanatory variables (if we include this assumption to the previous 4 we call them "Gauss-Markov Assumptions)

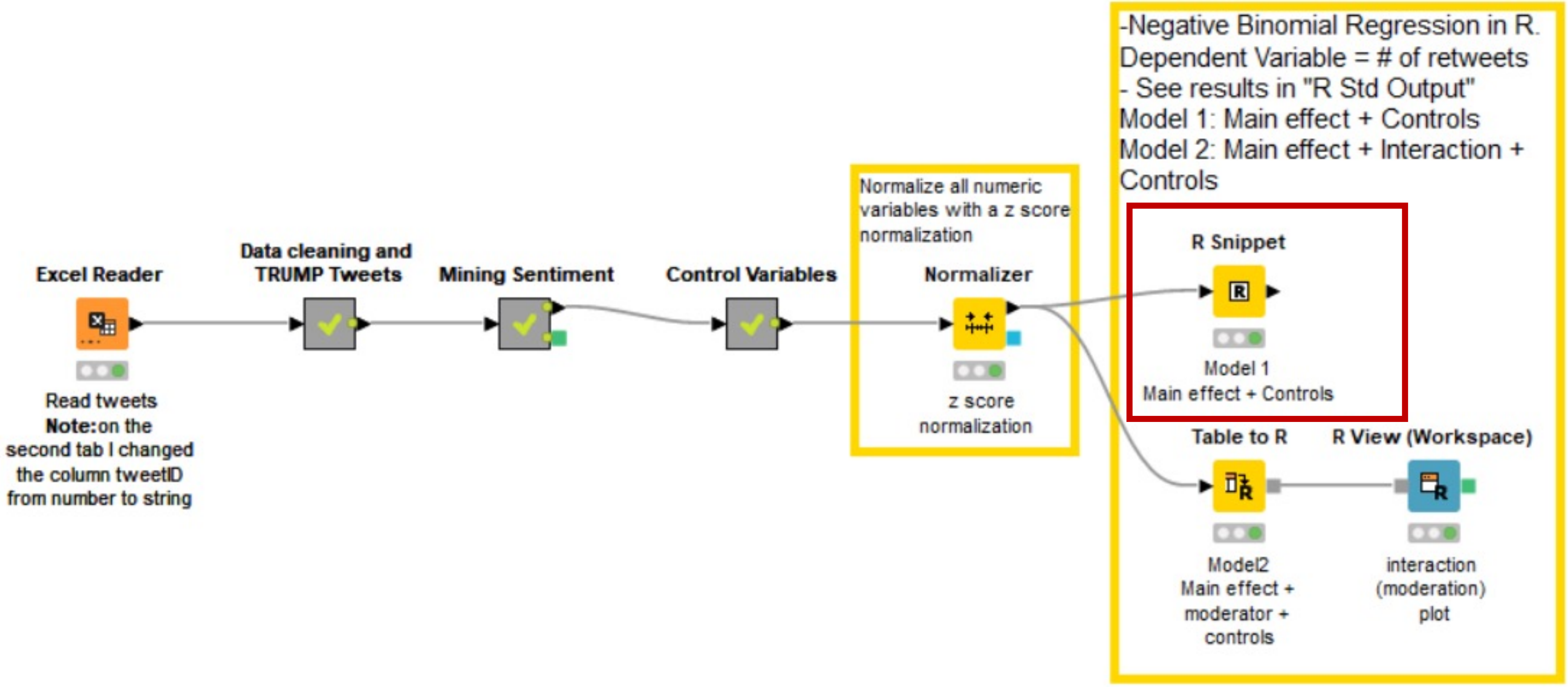
*Independent and dependent variables are metric (avoid nominal, counts; ratings can be used)

*These assumptions might vary when working with time series or panel data

Negative Binomial Regression

- Negative binomial regression is for modeling over dispersed (standard deviation greater than 2 times the mean) count variables.
- It is a non-linear type of model that uses maximum likelihood (ML; based on probability theory) for its estimation.
- Thus, the possible values of Y are the nonnegative integers: 0, 1, 2, 3, and so on.
- Negative binomial regression is a generalization of Poisson regression which loosens the restrictive assumption that the variance is equal to the mean made by the Poisson model

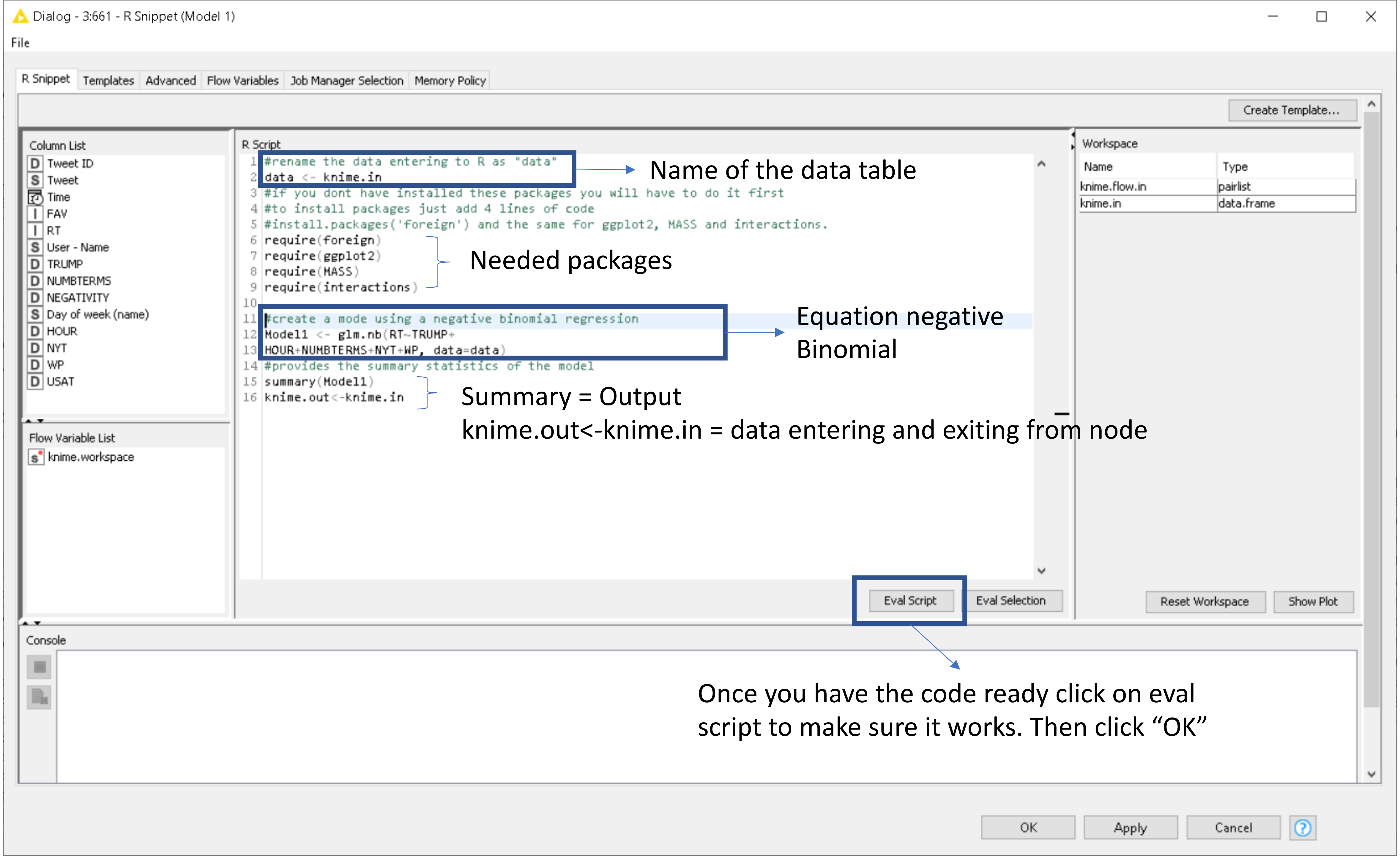
H1 = News media tweets containing information about Donald Trump will be more shared than tweets non related with Donald Trump



Model 1: Specification

$$Retweet_i = \exp(\propto + \beta_1 * TRUMP_i + \beta_2 * HOUR_i + \beta_3 * LENGTH_i + \beta_4 * NYT_i + \beta_5 * USATODAY_i + \mu_i)$$

- Dependent Variable = Retweet
- Main Effect (independent variable) = TRUMP
- Control Variables (to avoid omitted variables) = HOUR, LENGTH. Which others could be there?
- FIXED EFFECTS = NEWS MEDIA SOURCE (DUMMY VARIABLES). Which is the baseline of dummy variables?
- ERROR TERM = μ


Dialog - 3:661 - R Snippet (Model 1)
File
R Snippet
Templates
Advanced
Flow Variables
Job Manager Selection
Memory Policy
Create Template...
Column List
Tweet ID
Tweet
Time
FAV
RT
User - Name
TRUMP
NUMBTERMS
NEGATIVITY
Day of week (name)
HOUR
NYT
WP
USAT
Flow Variable List
knime.workspace
R Script
#rename the data entering to R as "data"
data <- knime.in
#if you dont have installed these packages you will have to do it first
#to install packages just add 4 lines of code
#install.packages('foreign') and the same for ggplot2, MASS and interactions.
require(foreign)
require(ggplot2)
require(MASS)
require(interactions)
#create a mode using a negative binomial regression
Model1 <- glm.nb(RT~TRUMP+
HOUR+NUMBTERMS+NYT+WP, data=data)
#provides the summary statistics of the model
summary(Model1)
knime.out<-knime.in
Name of the data table
Needed packages
Equation negative Binomial
Summary = Output
knime.out<-knime.in = data entering and exiting from node
Workspace
Name
Type
knime.flow.in
pairlist
knime.in
data.frame
Eval Script
Eval Selection
Reset Workspace
Show Plot
Console
Once you have the code ready click on eval script to make sure it works. Then click "OK"
OK
Apply
Cancel

Results Model 1(dependent variable: Retweets)

R Std Output - 3:661 - R Snippet (Model 1)

File

```
Call:
glm.nb(formula = RT ~ TRUMP + HOUR + NUMBTERMS + NYT + WP, data = data,
    init.theta = 0.6786637366, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8165  -1.0485  -0.6747  -0.1810  11.7599

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.79483    0.02522 190.117  < 2e-16 ***
TRUMP        0.24016    0.02548   9.427  < 2e-16 ***
HOUR        -0.19800    0.02550  -7.766 8.13e-15 ***
NUMBTERMS   -0.09451    0.03411  -2.771   0.0056 **
NYT          0.98638    0.03091  31.907  < 2e-16 ***
WP           0.74594    0.03412  21.863  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6787) family taken to be 1)

    Null deviance: 3837.7  on 2336  degrees of freedom
Residual deviance: 2801.0  on 2331  degrees of freedom
AIC: 26868

Number of Fisher Scoring iterations: 1


              Theta:  0.6787
          Std. Err.:  0.0171

 2 x log-likelihood:  -26853.7500
```

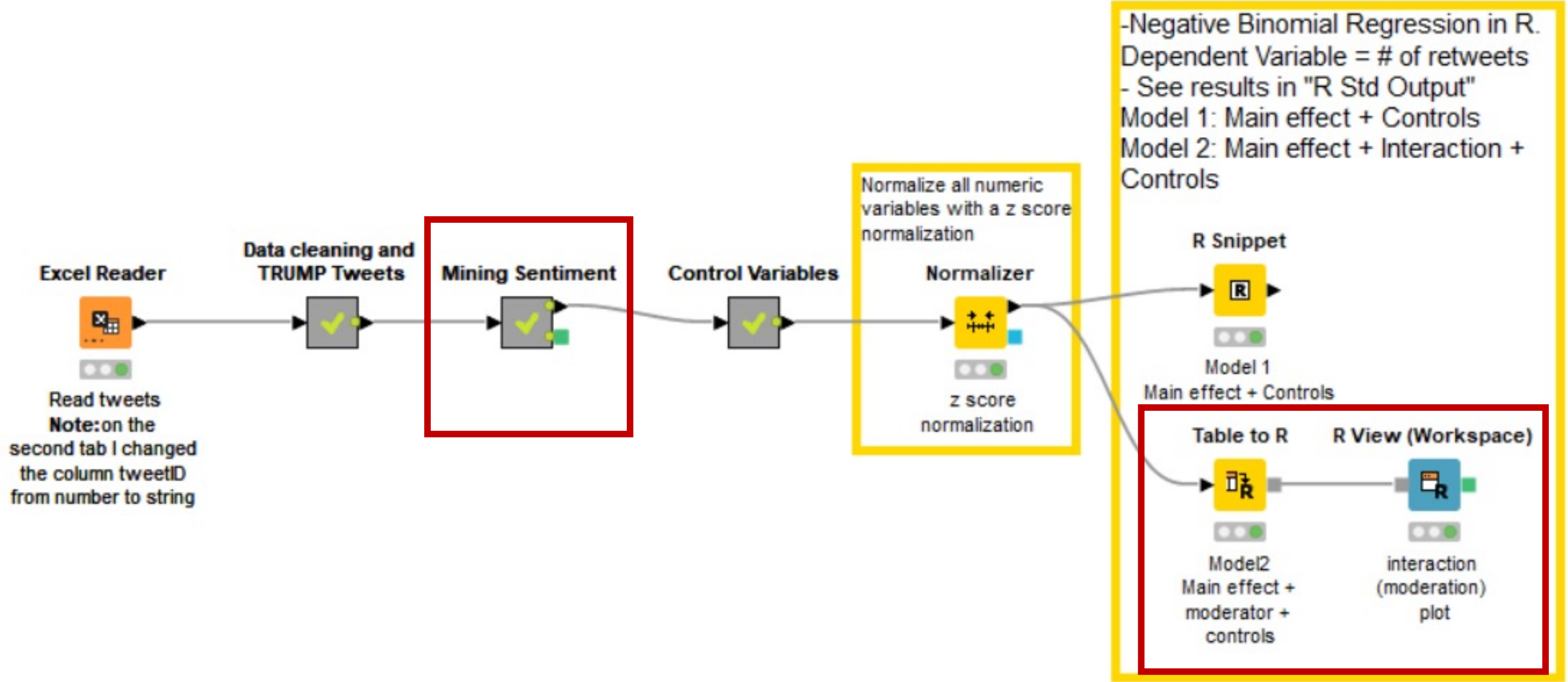

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.79483    0.02522 190.117  < 2e-16 ***
TRUMP        0.24016    0.02548   9.427  < 2e-16 ***
HOUR        -0.19800    0.02550  -7.766 8.13e-15 ***
NUMBTERMS   -0.09451    0.03411  -2.771   0.0056 **
NYT          0.98638    0.03091  31.907  < 2e-16 ***
WP           0.74594    0.03412  21.863  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6787) family taken to be 1)

    Null deviance: 3837.7  on 2336  degrees of freedom
Residual deviance: 2801.0  on 2331  degrees of freedom
AIC: 26868
```


*Standard for statistical significance: P value <0.05

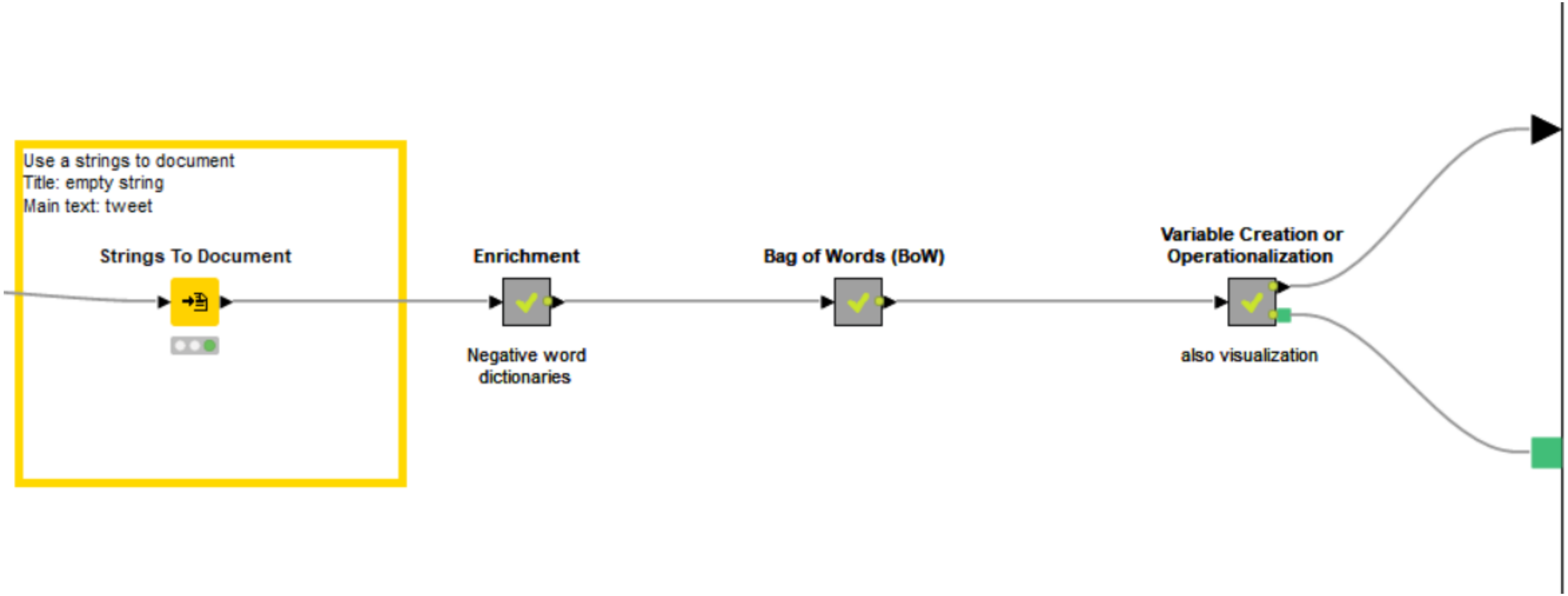
Interpretation?

H2 = Greater negative tone in news media tweets will result in greater consumer sharing

H3 = Greater negative tone in news media tweets will **weaken** the negative effect of TRUMP news on consumer sharing



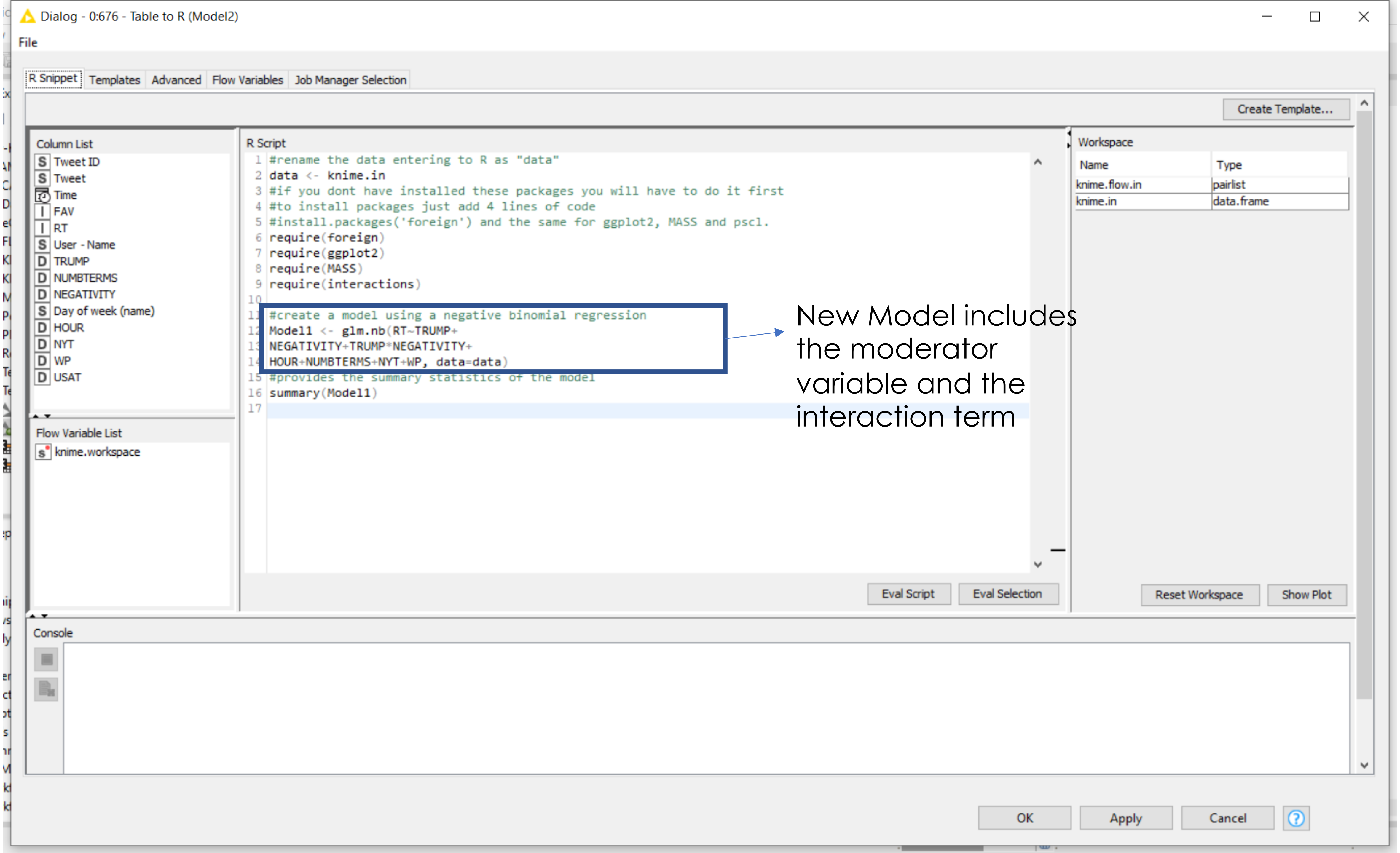
Metanode: Mining Sentiment (Negativity) Dictionary Method for Variable Operationalization



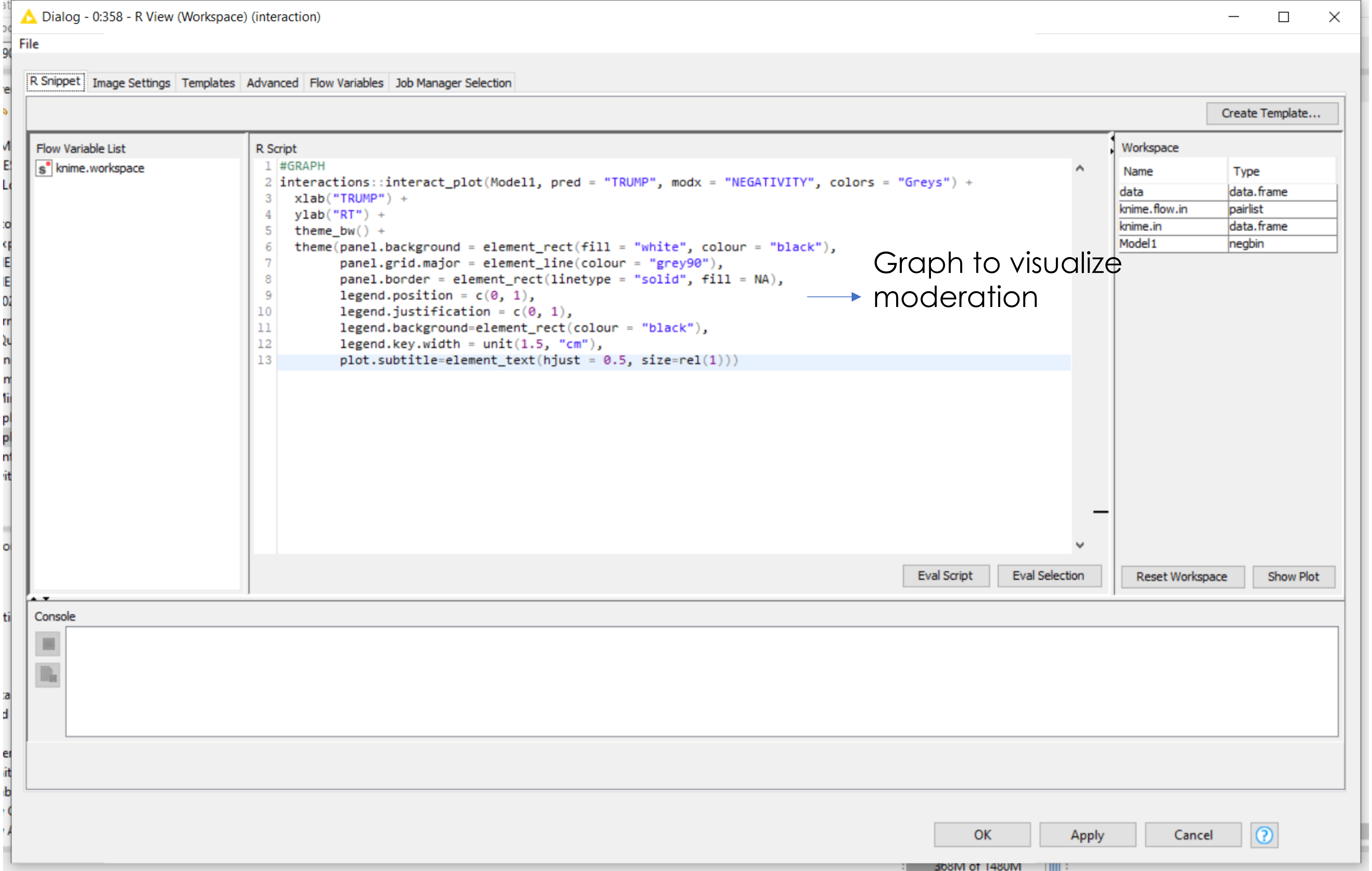
Model 2: Specification

$$Retweet_i = \exp\left(\propto + \beta_1 * TRUMP_i + \overbrace{\boldsymbol{\beta_2 * NEGATIVITY_i}}^{\text{Moderator Variable}} + \boldsymbol{\beta_3} * \overbrace{\boldsymbol{TRUMP_i * NEGATIVITY_i}}^{\text{Interaction Term}} + \beta_4 * HOUR_i + \beta_5 * LENGTH_i + \beta_6 * NYT_i + \beta_7 * USATODAY_i + \mu_i\right)$$

- Dependent Variable = Retweet
- Main Effect (independent variable) = TRUMP
- **Moderator/Interaction: NEGATIVITY**
- Control Variables (to avoid omitted variables) = HOUR, LENGTH. Which others could be there?
- FIXED EFFECTS = NEWS MEDIA SOURCE (DUMMY VARIABLES). Which is the baseline of dummy variables?
- ERROR TERM = μ



New Model includes the moderator variable and the interaction term


Dialog - 0:358 - R View (Workspace) (interaction)
File
R Snippet
Image Settings
Templates
Advanced
Flow Variables
Job Manager Selection
Create Template...
Flow Variable List
knime.workspace
R Script
#GRAPH
interactions::interact_plot(Model1, pred = "TRUMP", modx = "NEGATIVITY", colors = "Greys") +
xlab("TRUMP") +
ylab("RT") +
theme_bw() +
theme(panel.background = element_rect(fill = "white", colour = "black"),
panel.grid.major = element_line(colour = "grey90"),
panel.border = element_rect(linetype = "solid", fill = NA),
legend.position = c(0, 1),
legend.justification = c(0, 1),
legend.background=element_rect(colour = "black"),
legend.key.width = unit(1.5, "cm"),
plot.subtitle=element_text(hjust = 0.5, size=rel(1)))
Graph to visualize moderation
Workspace
Name
Type
data
data.frame
knime.flow.in
pairlist
knime.in
data.frame
Model1
negbin
Eval Script
Eval Selection
Reset Workspace
Show Plot
Console
OK
Apply
Cancel

Results Model 2(dependent variable: Retweets)

R Std Output - 3:669 - R Snippet (Model2)

File

```
Call:
glm.nb(formula = RT ~ TRUMP + NEGATIVITY + TRUMP * NEGATIVITY +
    HOUR + NUMBTERMS + NYT + WP, data = data, init.theta = 0.6827072274,
    link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8187  -1.0427  -0.6810  -0.1781  12.5455

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        4.78678    0.02516 190.286  < 2e-16 ***
TRUMP              0.24079    0.02543   9.469  < 2e-16 ***
NEGATIVITY         0.04623    0.02576   1.795 0.072692 .
HOUR              -0.18257    0.02543  -7.178 7.06e-13 ***
NUMBTERMS         -0.10051    0.03401  -2.955 0.003127 **
NYT                0.97975    0.03083  31.784  < 2e-16 ***
WP                 0.72041    0.03420  21.062  < 2e-16 ***
TRUMP:NEGATIVITY  -0.09310    0.02828  -3.292 0.000995 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6827) family taken to be 1)

    Null deviance: 3860.2  on 2336  degrees of freedom
Residual deviance: 2798.8  on 2329  degrees of freedom
AIC: 26853

Number of Fisher Scoring iterations: 1


              Theta:  0.6827
          Std. Err.:  0.0172

 2 x log-likelihood:  -26835.1780
```

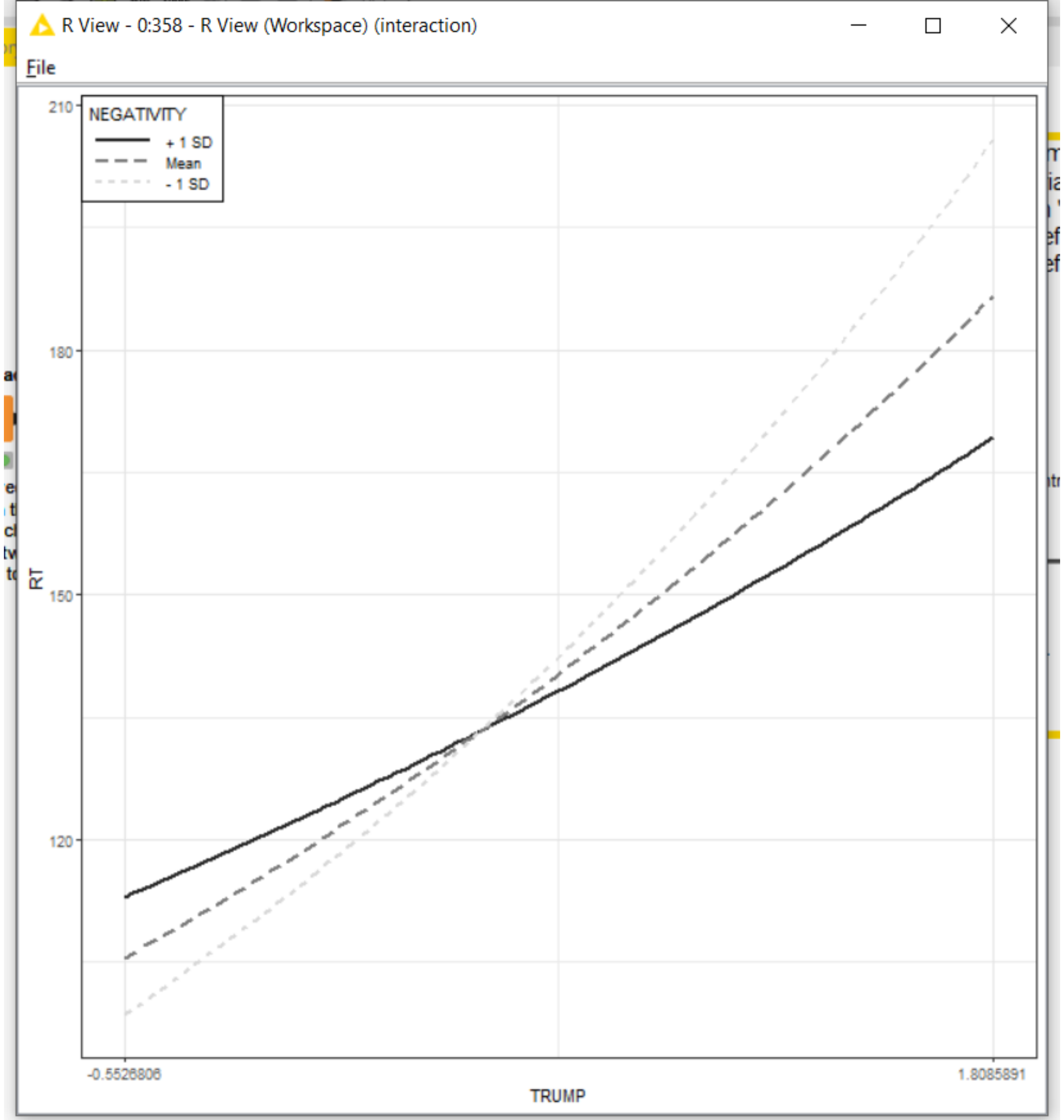
```
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       4.79233    0.02519 190.255  < 2e-16 ***
TRUMP             0.24151    0.02544   9.492  < 2e-16 ***
NEGATIVITY        0.02935    0.02542   1.154  0.24832
HOUR             -0.19077    0.02548  -7.488    7e-14 ***
NUMBTERMS        -0.10375    0.03411  -3.041  0.00236 **
NYT               0.98518    0.03088  31.902  < 2e-16 ***
WP                0.73597    0.03427  21.473  < 2e-16 ***
TRUMP:NEGATIVITY -0.07035    0.02637  -2.668  0.00763 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.6804) family taken to be 1)

    Null deviance: 3847.3  on 2336  degrees of freedom
Residual deviance: 2800.0  on 2329  degrees of freedom
AIC: 26864
```

*Standard for statistical significance: P value <0.05

Interpretation?

Interaction Plot

- Interpretation?



Model Comparison. DV = Retweets

Coefficients	MODEL 1	MODEL 2
(Intercept)	4.794**	4.792**
TRUMP	0.240**	0.241**
NEGATIVITY		0.029
HOUR	-0.198**	-0.190**
NUMBTERMS	-0.094**	-0.103**
NYT	0.986**	0.985**
WP	0.745**	0.735**
TRUMP:NEGATIVITY		-0.070**
AIC	26868	26864
N	2337	2337

- H1 = Supported
- H2 = Not Supported
- H3 = Supported
- Model Fit ?
- Improved from model 1 -> 2

Conclusions

The example demonstrated the most important steps in the explanatory analytics process, combining 1)managerial problem, 2)conceptualization and theory, 3)measurement and 4)model development.

1. Conceptualization
2. Hypothesis & theory
3. Model development
4. Measurement (Text Mining)
5. Estimation (Negative Binomial Regression
6. Hypothesis Testing
7. Findings & Conclusion

Beyond this simple exercise

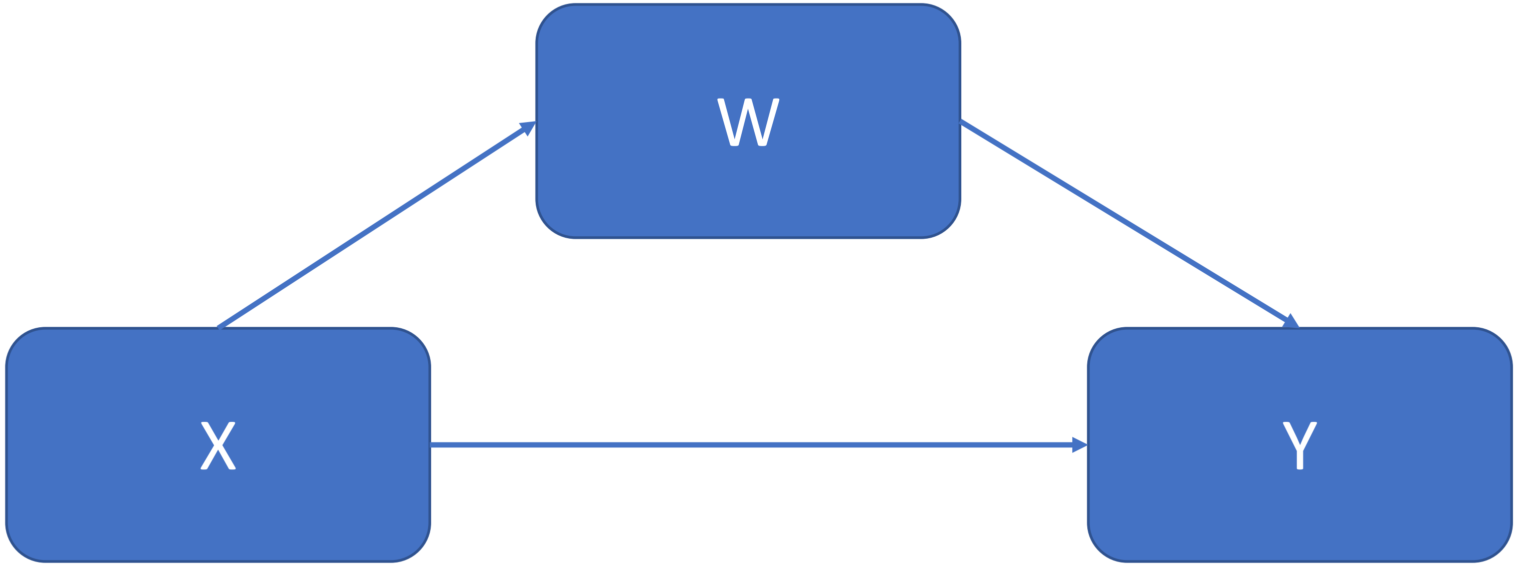
- Endogeneity Concerns
  - Selection Bias
  - Dynamic models (control functions)
- Panel Data Analysis & Time Series
- Other types of regression

References

- Book: Leeflang P., Wieringa J., Bijmolt T., and Pauwels K. (2015), "Modeling Markets, Analyzing Marketing Phenomena and Improving Marketing Decision Making." Springer, NY.
- Lindgreen, Adam, C. Anthony Di Benedetto, Roderick J. Brodie, and Elina Jaakkola. "How to develop great conceptual frameworks for business-to-business marketing." (2020).
- Shields, Patricia M., and Nandhini Rangarajan. A playbook for research methods: Integrating conceptual frameworks and project management. New Forums Press, 2013.
- Wooldridge, Jeffrey M. Introductory econometrics: A modern approach. Nelson Education, 2016.

Thanks

Mediation Effect



- Independent Variable (X)
- Dependent Variable (Y)
- Mediator Variable (M)

Mediation Effect

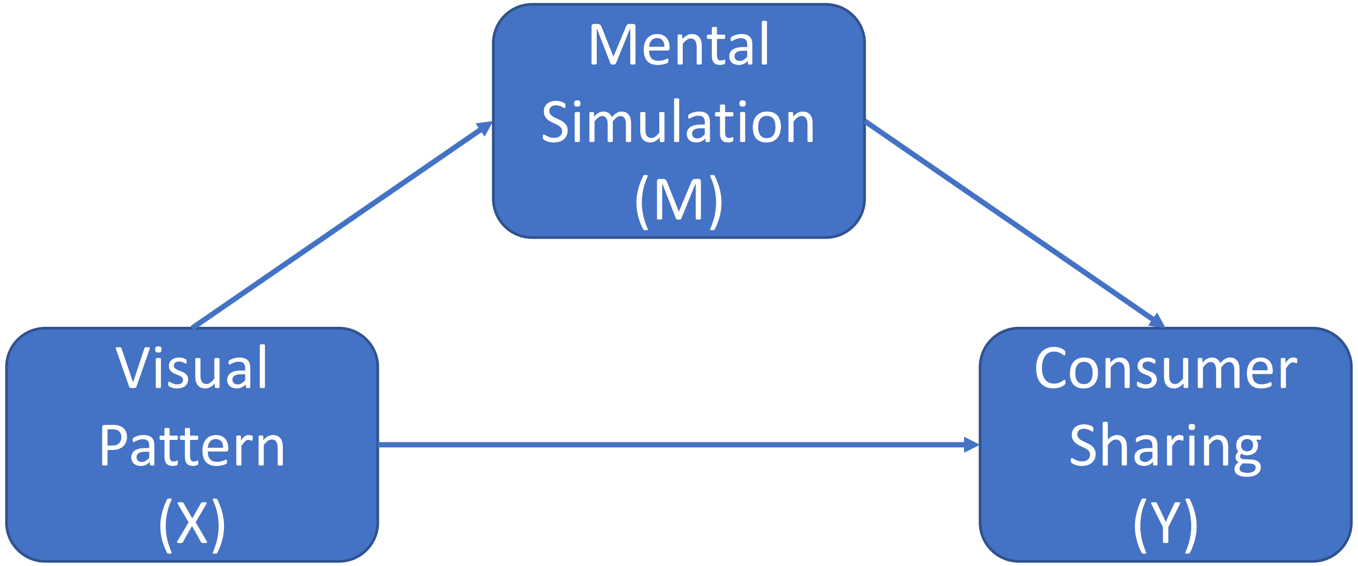
- Mediation implies a situation where the effect of the independent variable on the dependent variable can best be explained using a third mediator variable which is caused by the independent variable and is itself a cause for the dependent variable
- The mediator explains "how" (or some people says "why") a DV and IV are related
- A mediator is a variable that intervenes/intermediates in the relation between an independent variable and an outcome
- That is to say instead of X causing Y directly, X is causing the mediator M, and M is in turn causing Y. The causal relationship between X and Y in this case is said to be indirect.

Mediation Effect

- How or why do regular visual patterns enhance consumer sharing in social media?
- For example, does the level of mental simulation resulting from the visual patter affect the consumer sharing?

Mediation Effect

- How or why do regular visual patterns enhance consumer sharing in social media?
- For example, does the level of mental simulation resulting from the visual patter affect the consumer sharing?



- Independent Variable (X)
- Dependent Variable (Y)
- Moderator Variable (W)