

# Research Overview

Vineet Kumar

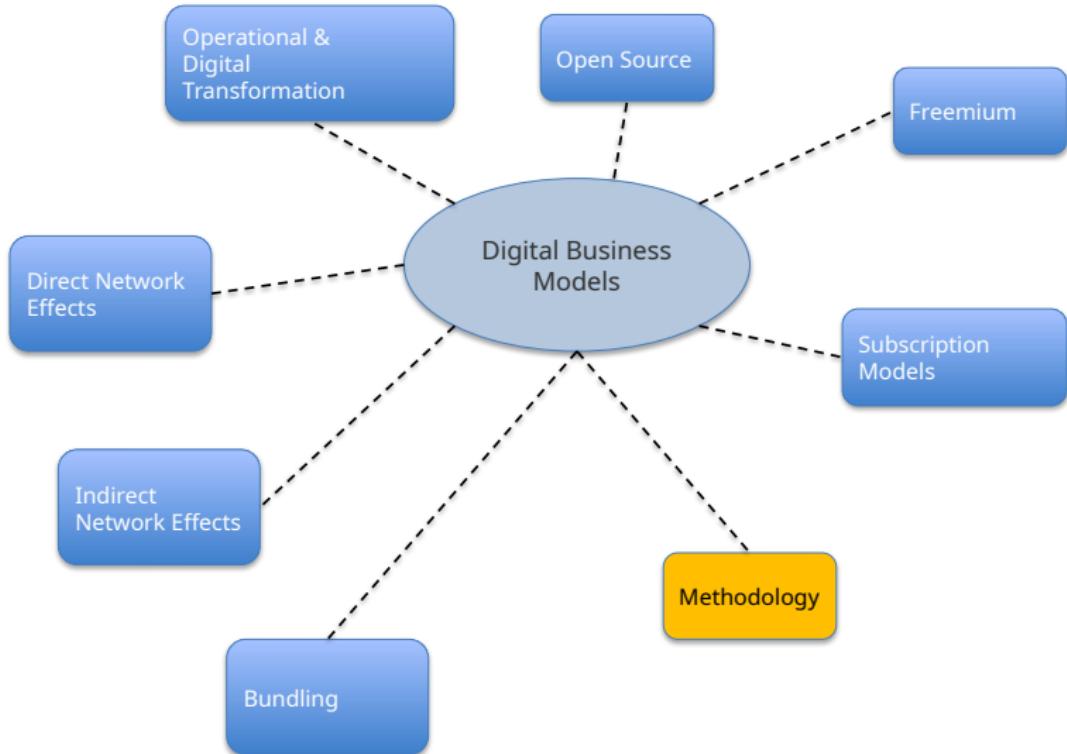
Yale School of Management

Presenting at:

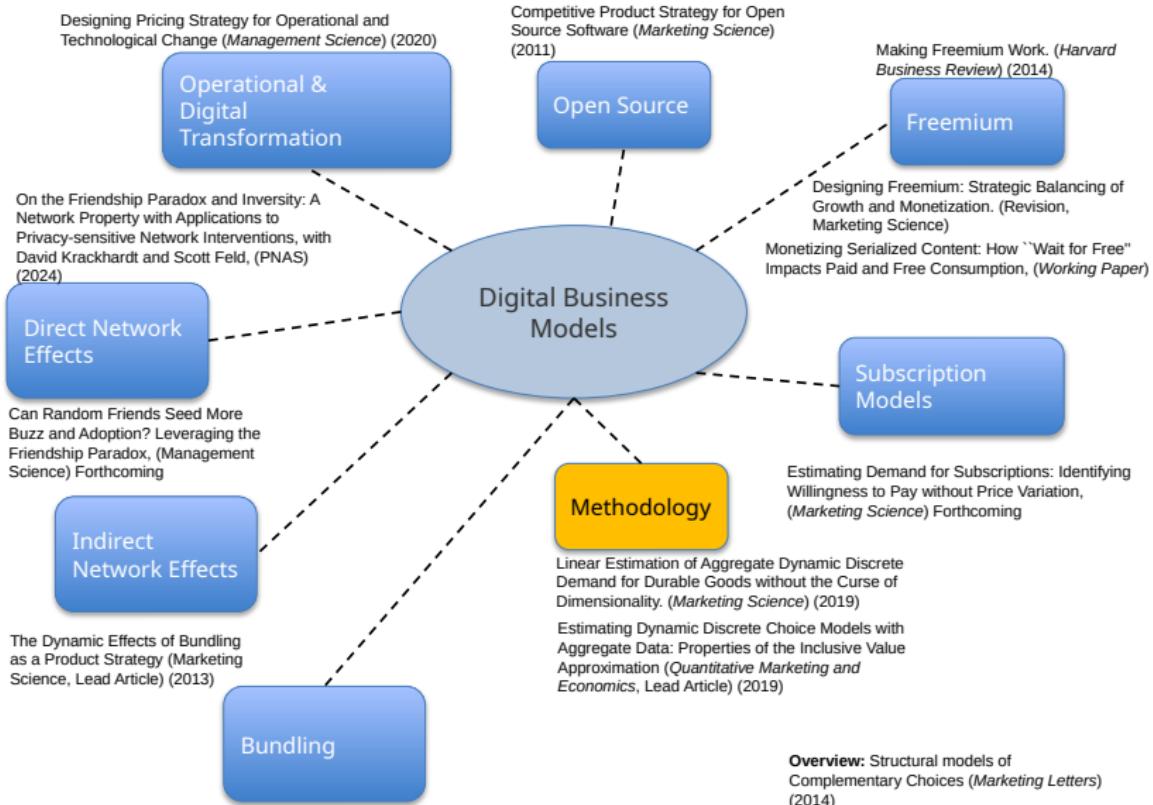
*Singapore Management University  
Lee Kong Chian School of Business*

September 2025

# Research Overview



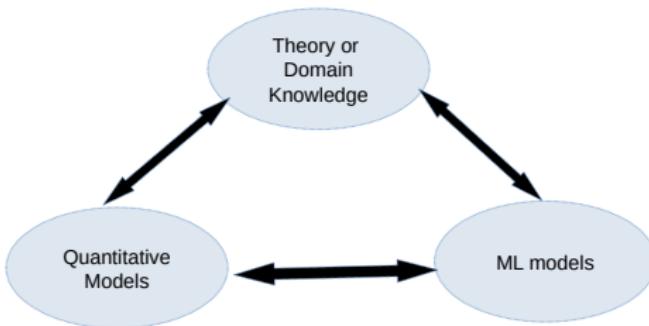
# Digital Economy – Business Models



**Overview:** Structural models of Complementary Choices (*Marketing Letters*) (2014)

# Role of Human Knowledge in Research

ML has typically been atheoretical



- One view of ML – advanced form of *statistical pattern matching*
  - Similar model (CNN) used both for detecting lung cancer (medicine) and for detecting stars (astronomy)

## My Take

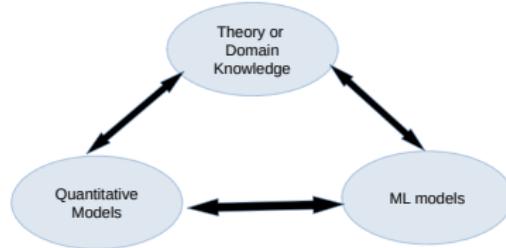
Our **domain knowledge (theory)** has a lot to add

# Role of Human Knowledge in Research

ML has typically been atheoretical

## Why add domain knowledge?

Can improve predictive *accuracy*, *explainability*, provide *guarantees*



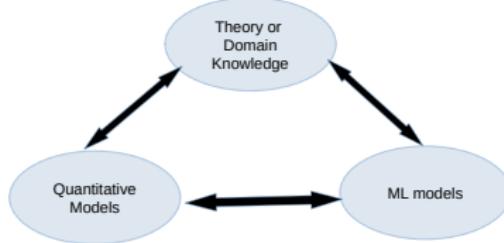
# Role of Human Knowledge in Research

ML has typically been atheoretical

## Why add domain knowledge?

Can improve predictive *accuracy, explainability, provide guarantees*

- A Theory-Based Explainable Deep Learning Architecture for Music Emotion. *Marketing Science* 44 (1), 196-219



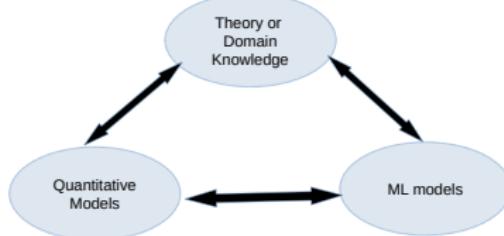
# Role of Human Knowledge in Research

ML has typically been atheoretical

## Why add domain knowledge?

Can improve predictive *accuracy, explainability, provide guarantees*

- A Theory-Based Explainable Deep Learning Architecture for Music Emotion. *Marketing Science* 44 (1), 196-219
- Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis. *Journal of Marketing Research* 62.3 (2025): 405-428

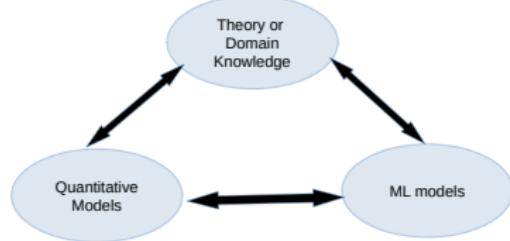


# Role of Human Knowledge in Research

ML has typically been atheoretical

## Why add domain knowledge?

Can improve predictive *accuracy, explainability, provide guarantees*



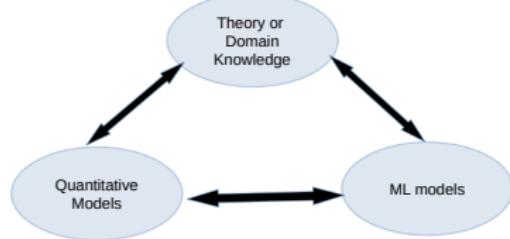
- A Theory-Based Explainable Deep Learning Architecture for Music Emotion. *Marketing Science* 44 (1), 196-219
- Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis. *Journal of Marketing Research* 62.3 (2025): 405-428
- Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve. Forthcoming at *Marketing Science*

# Role of Human Knowledge in Research

ML has typically been atheoretical

## Why add domain knowledge?

Can improve predictive *accuracy, explainability, provide guarantees*



- A Theory-Based Explainable Deep Learning Architecture for Music Emotion. *Marketing Science* 44 (1), 196-219
- Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis. *Journal of Marketing Research* 62.3 (2025): 405-428
- Nonparametric Bandits Leveraging Informational Externalities to Learn the Demand Curve. Forthcoming at *Marketing Science*
- Market Structure Mapping with Visual Characteristics. (Research in progress)

# Generative Interpretable Visual Design

Sisodia, Burnap and Kumar

Presenting at:  
*Singapore Management University*  
*Lee Kong Chian School of Business*  
September 2025

# Visual (or aesthetic) design matters across many product categories . . .



**Cars**

# Visual (or aesthetic) design matters across many product categories . . .



**Cars**



**Fashion**

# Visual (or aesthetic) design matters across many product categories . . .



**Cars**



**Fashion**



**Furniture**

... even for mundane categories like yogurt



*"We worked hard to get the packaging right ... American yogurt has always been sold in containers with relatively narrow openings. In Europe yogurt containers are wider and squatter, and that's what I wanted for Chobani."*

*—Hamdi Ulukaya, Founder & CEO, Chobani*

# Consumer Preferences for Visual Design



# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

What about preferences in visual space?

Cannot do this because characteristics for visual design are unknown!

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

What about preferences in visual space?

# Demand Estimation: Big Picture

Goal:

Obtain consumer preferences for visual design (conjoint or market data)

Demand Estimation for Products in Differentiated Product Markets in Economics and Marketing

- Builds on foundation of Lancaster (1966), Kotler (1967)
- Products are bundles or collections of **characteristics**
- Preferences over products  $\implies$  Preferences over characteristics makes this problem feasible.
- Car is represented as: (mpg, horsepower, range) etc.

What about preferences in visual space?

Cannot do this because characteristics for visual design are unknown!

# What this research seeks to do

## Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover and extract characteristics for products*

# What this research seeks to do

## Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products
- *quantify* these characteristics

# What this research seeks to do

## Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products
- *quantify* these characteristics
- *generate visual designs that span the space of visual characteristics*

# What this research seeks to do

## Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products
- *quantify* these characteristics
- *generate* visual designs that span the space of visual characteristics

# What this research seeks to do

## Research Goals

Obtain **human-interpretable** visual characteristics (not outliers) directly from unstructured product image data:

- *automatically discover* and extract characteristics for products
- *quantify* these characteristics
- *generate* visual designs that span the space of visual characteristics



Hyundai: (3, 8, 5, 9) compared to BMW: (1, 3, 10, 1)

# Research Goals

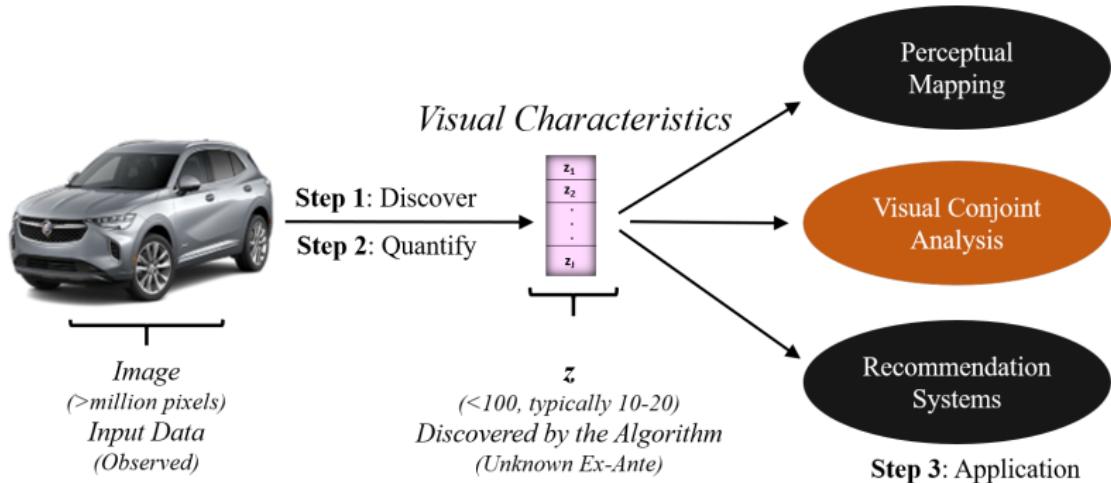


Hyundai: (3, 8, 5, 9) compared to BMW: (1, 3, 10, 1)

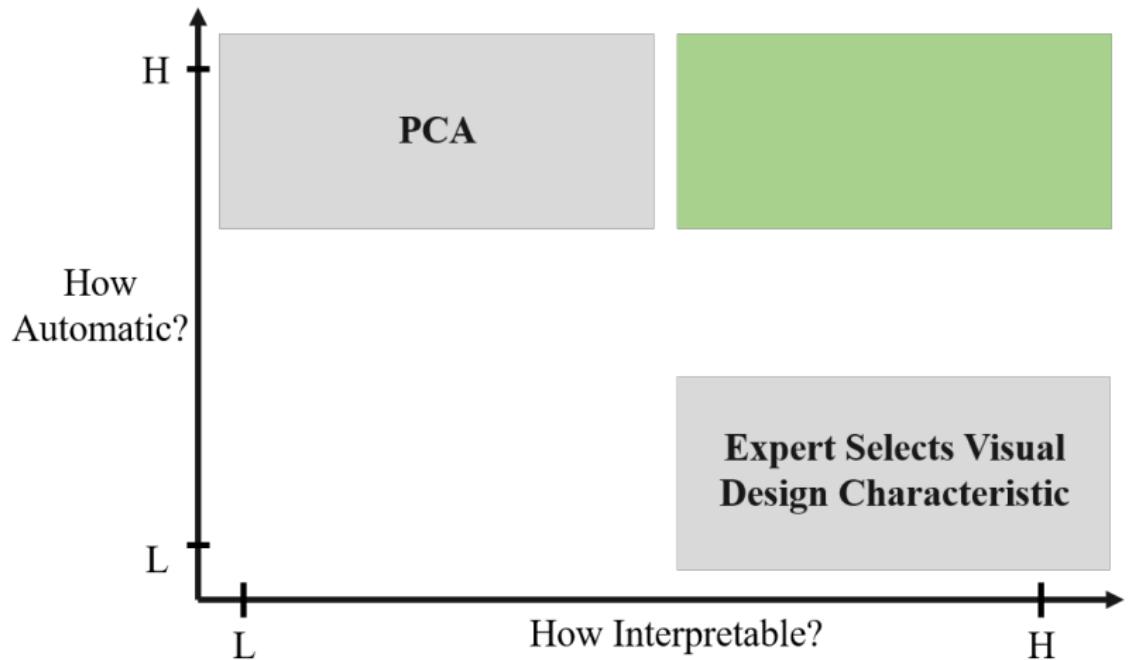
Several questions come to mind:

- What does the first number represent? Does 3 mean something different from 1?
- Can humans interpret these numeric values?
- What domain knowledge does the model need to have?

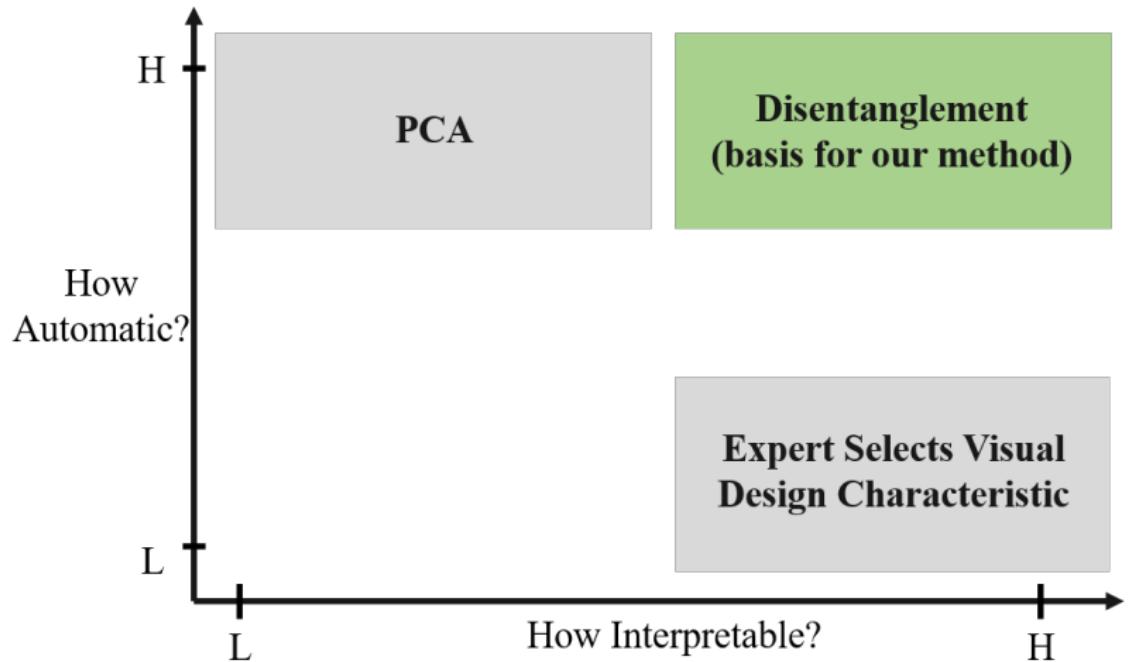
# Why Visual Characteristics?



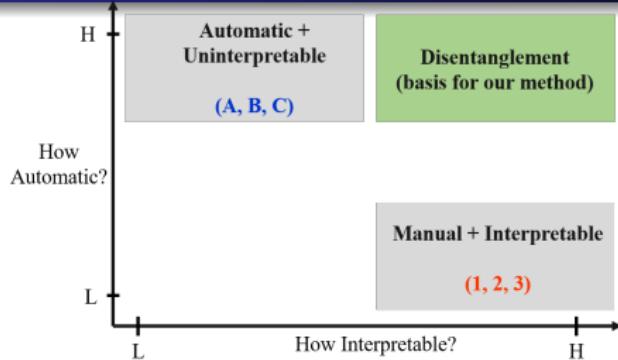
# Modeling Visual Characteristics: A comparison of methods



# Modeling Visual Characteristics: A comparison of methods



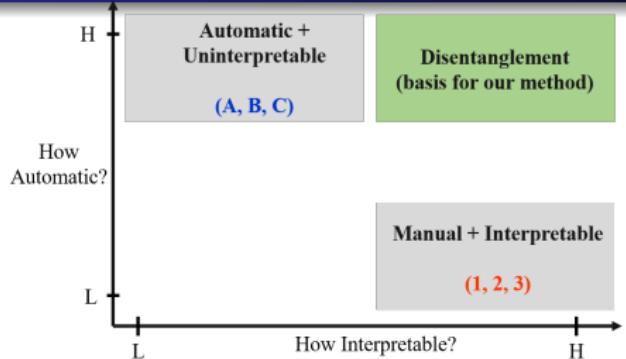
# Modeling Visual Characteristics: A comparison of methods



## Automatic + Uninterpretable

- A - Bajari, P. L. et al. (2021) : Hedonic prices and quality adjusted price indices powered by AI, *CENMAP working paper*
- B - Law, S., et al. (2019) : Take a look around: using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*
- C - Aubry, S., et al. (2019) : Machine learning, human experts, and the valuation of real assets. *CFS Working Paper Series*

# Modeling Visual Characteristics: A comparison of methods



## Automatic + Uninterpretable

- A - Bajari, P. L. et al. (2021) : Hedonic prices and quality adjusted price indices powered by AI, *CENMAP working paper*
- B - Law, S., et al. (2019) : Take a look around: using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology (TIST)*
- C - Aubry, S., et al. (2019) : Machine learning, human experts, and the valuation of real assets. *CFS Working Paper Series*

## Manual + Interpretable

- 1 - Zhang, M. et al. (2022) : Can consumer-posted photos serve as a leading indicator of restaurant survival? Evidence from yelp. *Management Science*
- 2 - Liu, Y., et al. (2017) : The effects of products' aesthetic design on demand and marketing-mix effectiveness: The role of segment prototypicality and brand consistency. *Journal of Marketing*
- 3 - Zhang, S., et al. (2021) : What makes a good image? Airbnb demand analytics leveraging interpretable image features. *Management Science*

# What is disentanglement?

Bengio et al (2013)

*“A disentangled representation can be defined as one where **single latent units** are sensitive to changes in **single generative factors**, while being relatively invariant to changes in other factors”*

# What is disentanglement?

Bengio et al (2013)

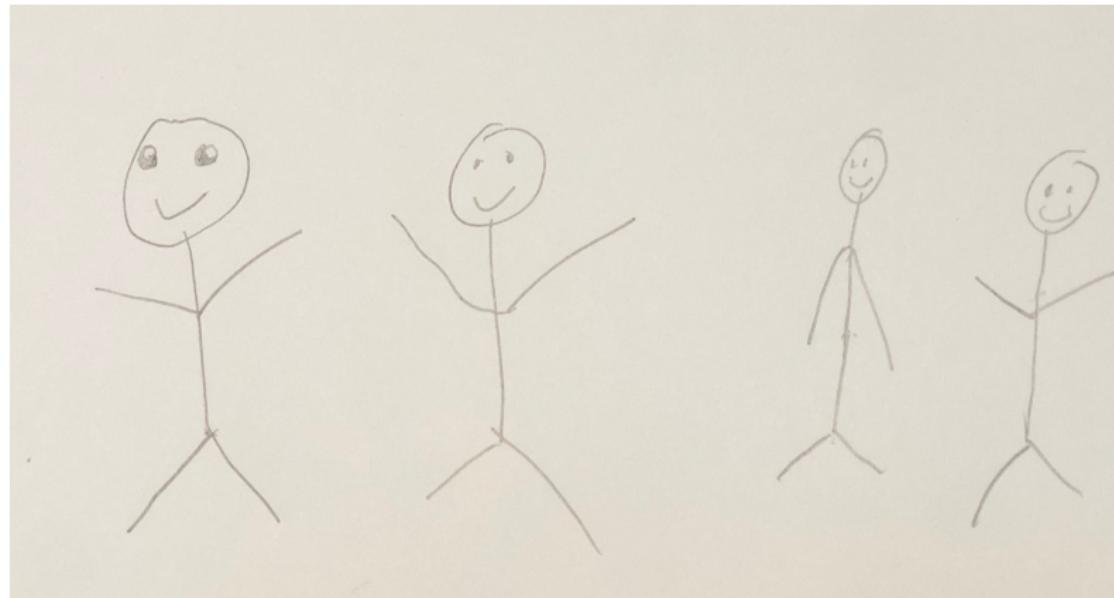
*"A disentangled representation can be defined as one where **single latent units** are sensitive to changes in **single generative factors**, while being relatively invariant to changes in other factors"*

- Latent Units ( $\mathbf{z}$ ): Dimensions in the model's latent space
- Generative factors ( $\mathbf{c}$ ): Human-interpretable true characteristics

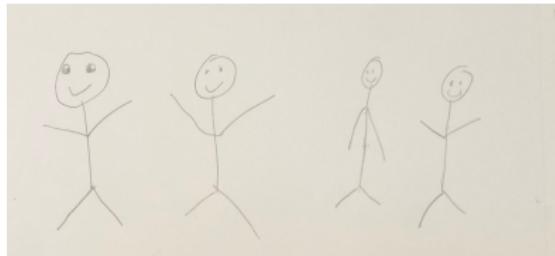
Idea: Reality or Data generating process is compositional based on generative factors.

# What is disentanglement?

Stick



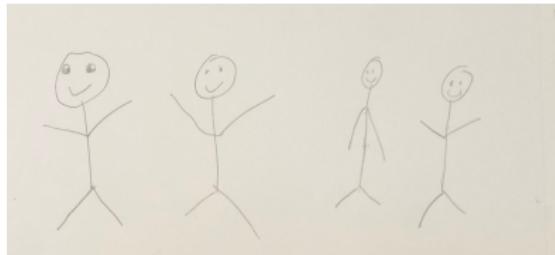
# What is disentanglement?



Bengio et al (2013)

*"A disentangled representation can be defined as one where **single latent units** are sensitive to changes in **single generative factors**, while being relatively invariant to changes in other factors"*

# What is disentanglement?

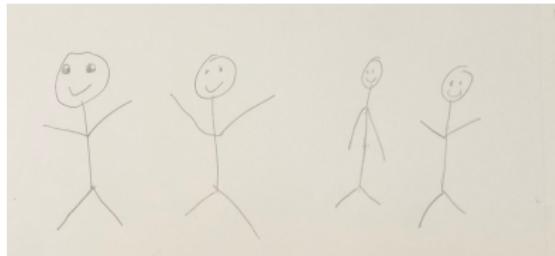


Bengio et al (2013)

*"A disentangled representation can be defined as one where **single latent units** are sensitive to changes in **single generative factors**, while being relatively invariant to changes in other factors"*

- Latent Units ( $\mathbf{z}$ ): What algorithm discovers – dimensions in the model's latent space
- Generative factors ( $\mathbf{c}$ ): Human-interpretable

# What is disentanglement?



Bengio et al (2013)

*"A disentangled representation can be defined as one where **single latent units** are sensitive to changes in **single generative factors**, while being relatively invariant to changes in other factors"*

- Latent Units ( $z$ ): What algorithm discovers – dimensions in the model's latent space
- Generative factors ( $c$ ): Human-interpretable

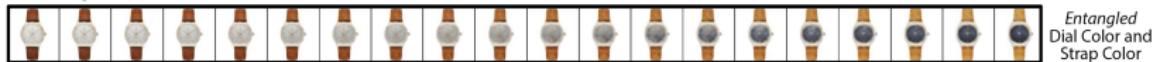
Goal: One to one mapping between  $z \Leftrightarrow c$

# Product Images and Parts of Watch



# Disentangled and Entangled Representations

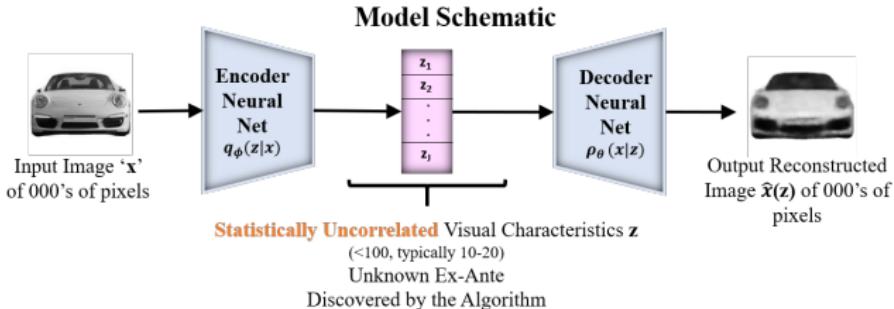
Example of *Entangled* Visual Characteristics



Example of *Disentangled* Visual Characteristics

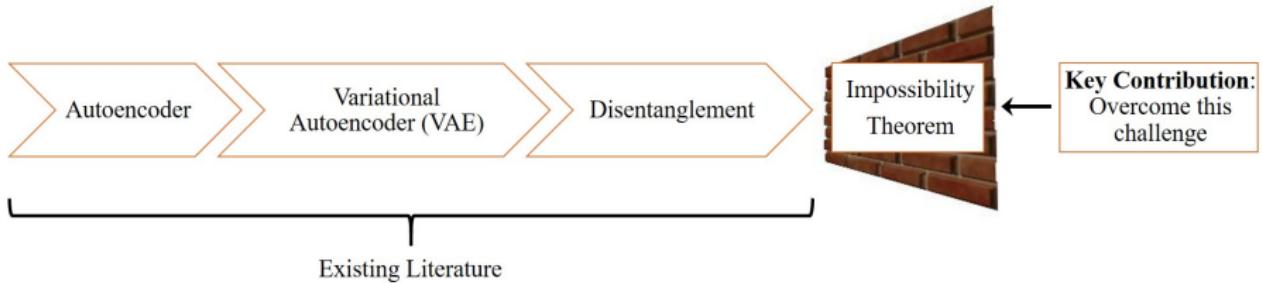


# Models in Existing Literature



Model	Goal
Autoencoder (AE)	Reconstruction accuracy
Variational Autoencoder (VAE)	... + structured latent space
Disentanglement	... + ... + statistically independent latent space

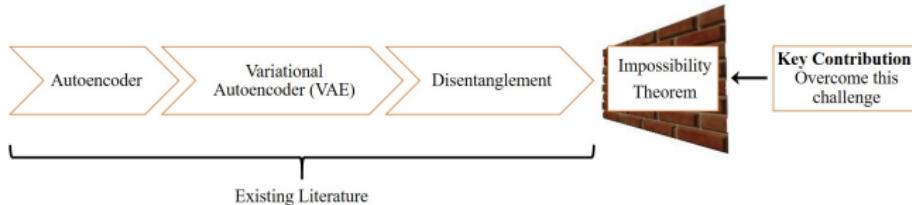
# Roadmap of Our Research Approach



## Contribution

We aim to overcome this impossibility theorem with a simple approach of using structured product characteristics.

# Impossibility Theorem



## Impossibility Theorem

Unsupervised (*i.e. only images*) learning of disentangled representations is *fundamentally impossible* except under certain restrictive conditions.<sup>a</sup>

<sup>a</sup>Locatello, Francesco, et al. "Challenging common assumptions in the unsupervised learning of disentangled representations." ICML. PMLR, 2019.

**Implication:** Every disentangled representation can have other *infinite* equivalent entangled representations.

# ML Approach to Impossibility Theorem

**Impossibility:** Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility

# ML Approach to Impossibility Theorem

**Impossibility:** Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility
- Need a supervisory signal

# ML Approach to Impossibility Theorem

**Impossibility:** Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility
- Need a supervisory signal
- **ML methods assume that ground truth is known by researchers**

# ML Approach to Impossibility Theorem

**Impossibility:** Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility
- Need a supervisory signal
- ML methods assume that ground truth is known by researchers
  - Human labeling

# ML Approach to Impossibility Theorem

**Impossibility:** Without Supervision, every disentangled representation can have other *infinite* equivalent entangled representations.

- ML researchers recognize the challenge of impossibility
- Need a supervisory signal
- ML methods assume that ground truth is known by researchers
  - Human labeling
- Can we use this approach to discover visual characteristics?

# Impossibility Theorem – Implications



*predicts*

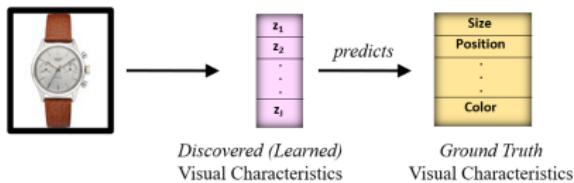
A horizontal black arrow pointing from right to left, indicating a flow or relationship between the learned features and the ground truth characteristics.

*Discovered (Learned)*  
Visual Characteristics

*Ground Truth*  
Visual Characteristics

# Impossibility Theorem – Implications

Common approach to ground truth in ML is to get humans to label<sup>1</sup>



## What's the Problem?

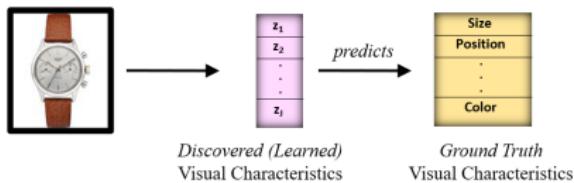
- Ground truth on visual characteristics is *unknown*.

<sup>1</sup>

Locatello, Francesco, et al. "Disentangling factors of variation using few labels." ICLR. 2020.

# Impossibility Theorem – Implications

Common approach to ground truth in ML is to get humans to label<sup>1</sup>



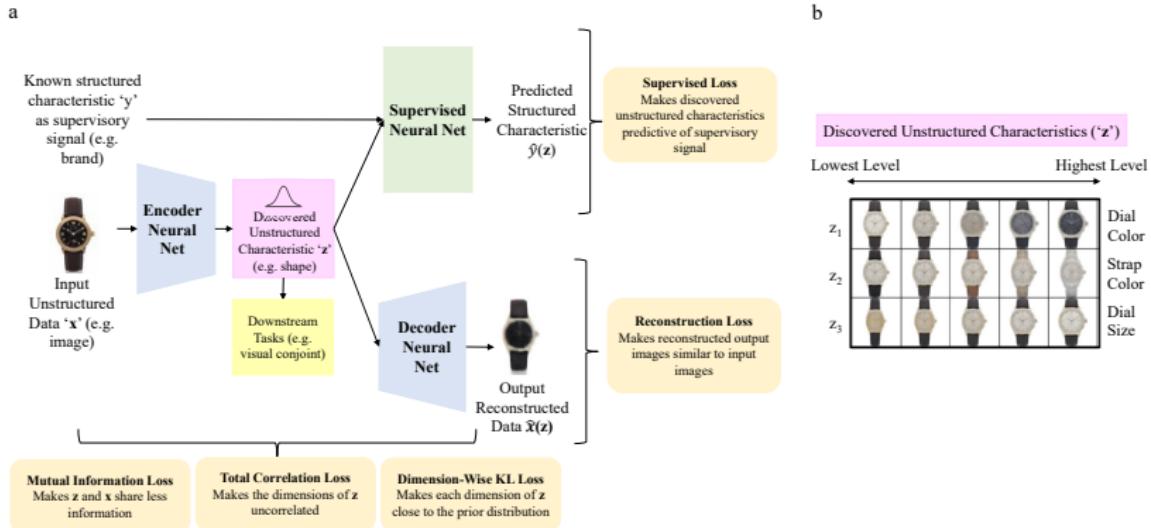
## What's the Problem?

- Ground truth on visual characteristics is *unknown*.
- Need to ensure humans understand what these labels are and *how to quantify them* for each image

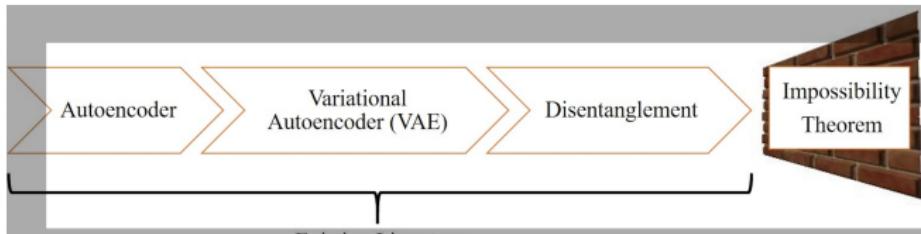
<sup>1</sup>

Locatello, Francesco, et al. "Disentangling factors of variation using few labels." ICLR. 2020.

# Schematic of Proposed Approach



# Contribution



- **Solution** without ground truth on visual characteristics:
- Leverage **structured product characteristics** to provide a supervisory signal for disentanglement

# Model

- $\phi$  encoder and  $\theta$  decoder parameters;  $\mathbf{x}$  are images and  $\mathbf{z}$  are visual characteristics
- Learn model parameters by minimizing loss  $L(\theta, \phi; \mathbf{x}, \mathbf{z})$  of

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} \\ + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

# Model

- $\phi$  encoder and  $\theta$  decoder parameters;  $\mathbf{x}$  are images and  $\mathbf{z}$  are visual characteristics
- Learn model parameters by minimizing loss  $L(\theta, \phi; \mathbf{x}, \mathbf{z})$  of

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

Loss Term	Why is this term included?
Reconstruction	Promotes accurate reconstruction of images
Mutual Information	Minimizes redundant information
<b>Total Correlation</b>	<b>Promotes statistical independence between visual characteristics</b>
Dimension-Wise KL	Penalizes deviations from a prior
<b>Supervised</b>	<b>Provides a signal to address the impossibility theorem</b>

# Model: Reconstruction Loss

- $\phi$  encoder and  $\theta$  decoder parameters;  $\mathbf{x}$  are images and  $\mathbf{z}$  are visual characteristics
- Learn model parameters by minimizing loss  $L(\theta, \phi; \mathbf{x}, \mathbf{z})$  of integrated model
- Reconstruction Loss (pixel level):

$$E_{q_\phi(z|x)}[\log p_\theta(x|z)]$$

- Given latent representation  $z$ , can decoder accurately obtain the  $x$  in the data?
- $p_\theta(x|z)$  high probability if reconstruction is good
- Encoder produces  $z$  probabilistically, so need to take expectations  $E_{q_\phi(z|x)}$
- Both encoder and decoder are learning here

# Model: Mutual Information Loss

- $\phi$  encoder and  $\theta$  decoder parameters;  $\mathbf{x}$  are images and  $\mathbf{z}$  are visual characteristics
- Learn model parameters by minimizing loss  $L(\theta, \phi; \mathbf{x}, \mathbf{z})$  of integrated model
- Mutual Information:

$$I_q(X; Z) = \mathbb{E}_{q(x,z)} \left[ \log \left( \frac{q(x,z)}{q(x)q(z)} \right) \right]$$

- Each input  $x$  produces a distinct  $z$ , retains lot of information

# Model

- $\phi$  encoder and  $\theta$  decoder parameters;  $\mathbf{x}$  are images and  $\mathbf{z}$  are visual characteristics
- Learn model parameters by minimizing loss  $L(\theta, \phi; \mathbf{x}, \mathbf{z})$  of

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

# Model

- $\phi$  encoder and  $\theta$  decoder parameters;  $\mathbf{x}$  are images and  $\mathbf{z}$  are visual characteristics
- Learn model parameters by minimizing loss  $L(\theta, \phi; \mathbf{x}, \mathbf{z})$  of

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

Loss Term	Why is this term included?
Reconstruction	Promotes accurate reconstruction of images
Mutual Information	Minimizes redundant information
<b>Total Correlation</b>	<b>Promotes statistical independence between visual characteristics</b>
Dimension-Wise KL	Penalizes deviations from a prior
<b>Supervised</b>	<b>Provides a signal to address the impossibility theorem</b>



# Model – Role of Supervised Loss

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} \\ + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

- Supervised Loss is used to predict signal from latent representation  $z$ :  $s = f(z)$

# Model – Role of Supervised Loss

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} \\ + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \underbrace{\delta P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

- Supervised Loss is used to predict signal from latent representation  $z$ :  $s = f(z)$
- Can use structured product characteristics as signals: brand, price, material etc.

# Model – Role of Supervised Loss

$$\begin{aligned} L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z}) &= \underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} \\ &\quad + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}} \end{aligned}$$

- Supervised Loss is used to predict signal from latent representation  $z$ :  $s = f(z)$
- Can use structured product characteristics as signals: brand, price, material etc.

# Model – Role of Supervised Loss

$$\underbrace{L(\theta, \phi, \mathbf{w}; \mathbf{x}, \mathbf{z})}_{\text{Total Loss}} = \underbrace{\mathbf{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})]}_{\text{Reconstruction Loss}} + \alpha \underbrace{I_q(\mathbf{z}, \mathbf{x})}_{\text{Mutual Information Loss}} + \beta \underbrace{KL \left[ q(\mathbf{z}) || \prod_{j=1}^J q(z_j) \right]}_{\text{Total Correlation Loss}} \\ + \gamma \underbrace{\sum_{j=1}^J KL \left[ q(z_j) || p(z_j) \right]}_{\text{Dimension-Wise KL Divergence Loss}} + \delta \underbrace{P(\hat{\mathbf{y}}(\mathbf{z}), \mathbf{y})}_{\text{Supervised Loss}}$$

- Supervised Loss is used to predict signal from latent representation  $z$ :  $s = f(z)$
- Can use structured product characteristics as signals: brand, price, material etc.

## Idea to Overcome Impossibility Theorem

If the supervisory signal is sufficiently correlated with visual characteristics, then it can help obtain the unique (true) disentangled representation

# Why might brand aid the disentanglement model?



## Brand as a Supervisory Signal

Idea: Brands have a specific “look” that can be correlated with visual appearance (and therefore visual characteristics)

# Evaluating Visual Characteristics

# Visual Characteristics: Human Interpretable?

- Are these visual characteristics human interpretable?

# Visual Characteristics: Human Interpretable?

- Are these visual characteristics human interpretable?



Starting from the image on the left, **what part of the watch changes the most** as you go from left to right? Carefully check both large and small visual aspects. Go through each part of the watch one by one before selecting any option. Refer to the above image to see parts of the watch.



Note: Images are low-quality on purpose

- |                                   |                                   |
|-----------------------------------|-----------------------------------|
| <input type="radio"/> Bezel       | <input type="radio"/> Hands       |
| <input type="radio"/> Crown       | <input type="radio"/> Hour Marker |
| <input type="radio"/> Date Window | <input type="radio"/> Lug         |
| <input type="radio"/> Dial        | <input type="radio"/> Strap       |

How is that part of the watch changing?

# Visual Characteristics: Quantification?

## Interpretability and Quantification

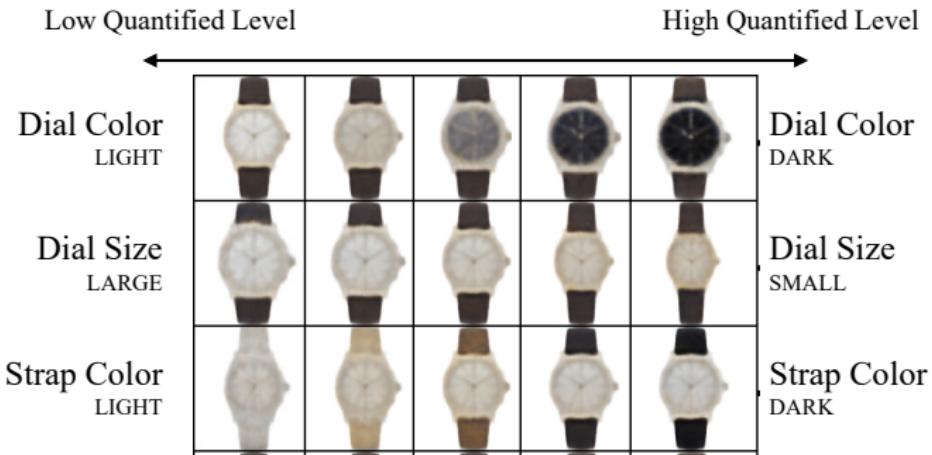
Visual characteristic	Interpretability Survey	Quantification Survey
Dial Size	76%	83%
Dial Color	80%	92%
Strap Color	88%	92%
Rim (Bezel) Color	79%	88%
Dial Shape	87%	68%
Knob (Crown) Size	70%	85%

# Discovered Visual characteristics

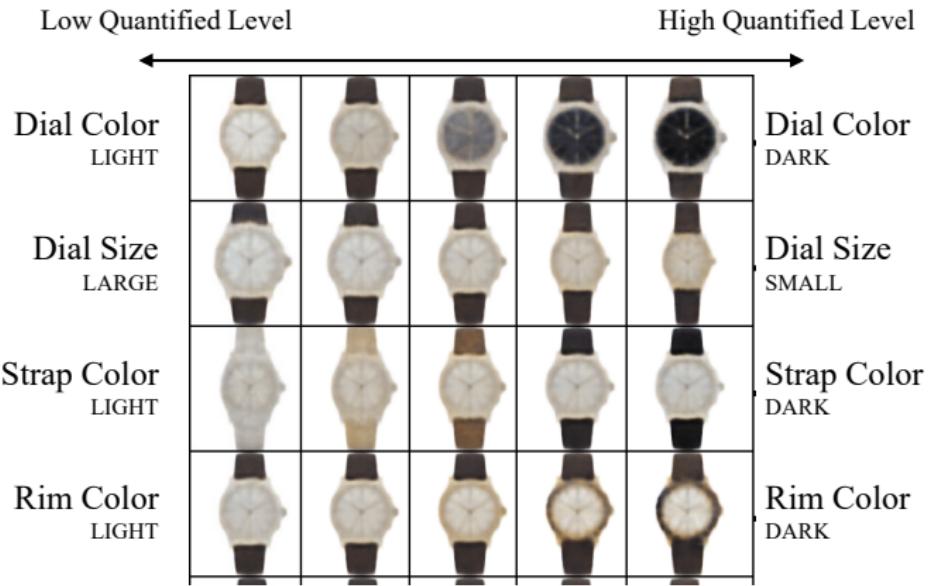
# Discovered Visual characteristics



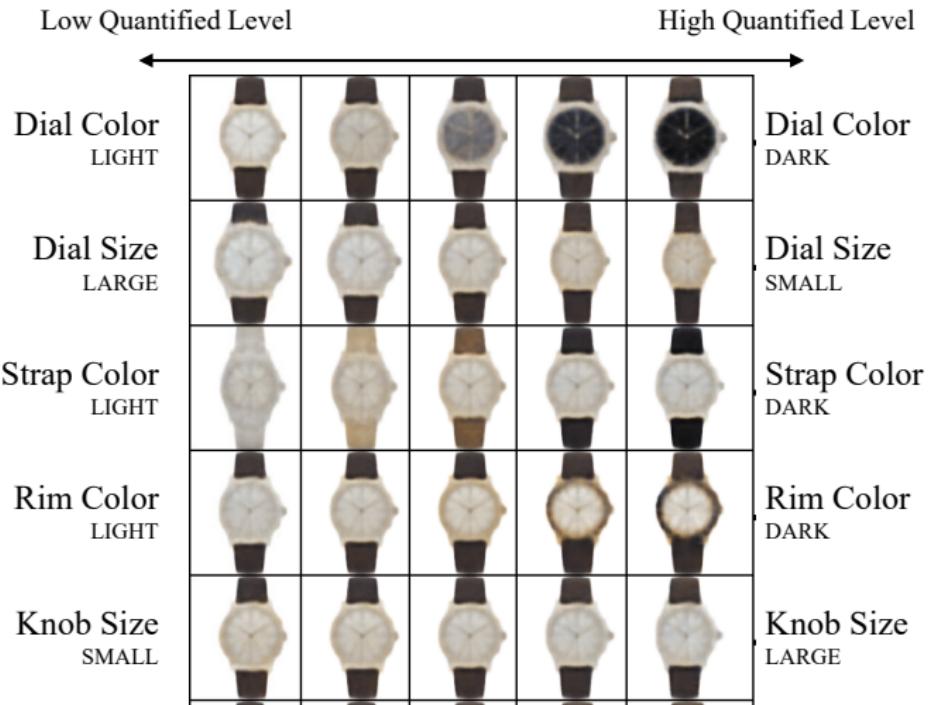
# Discovered Visual characteristics



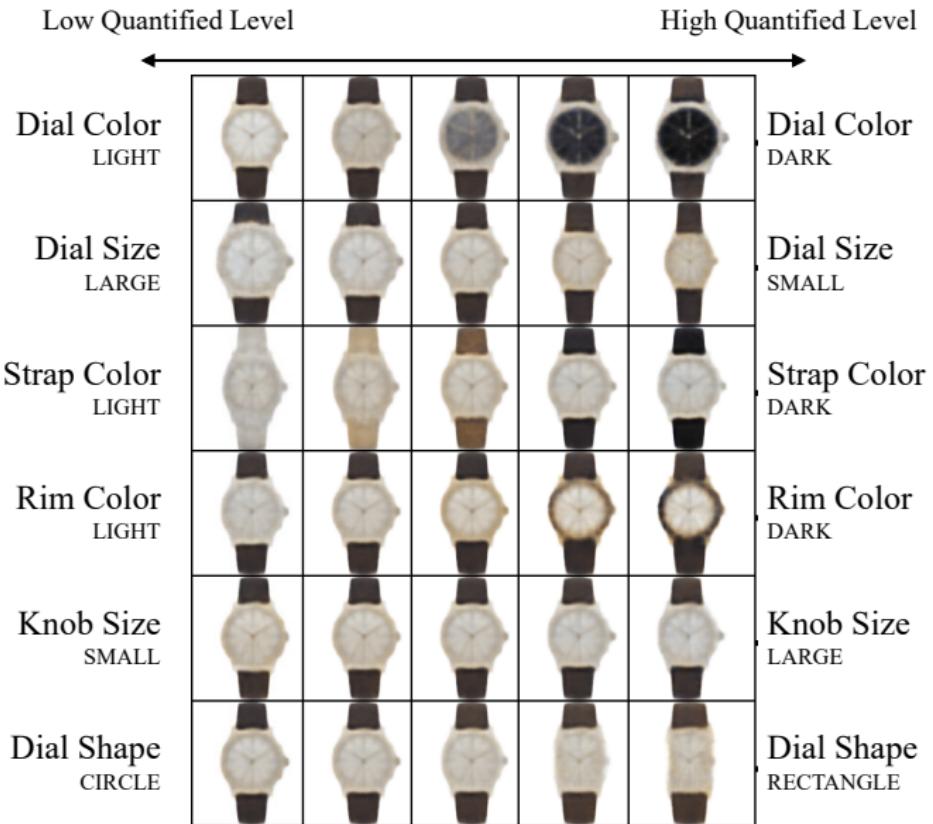
# Discovered Visual characteristics



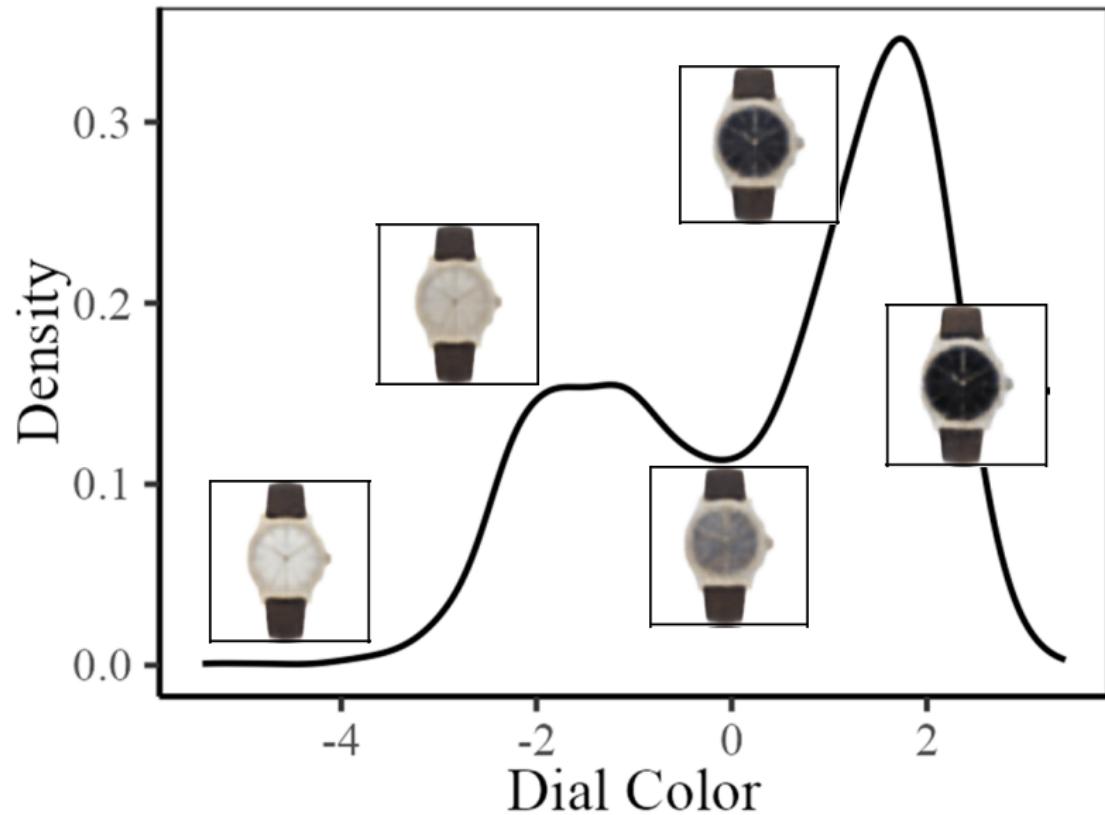
# Discovered Visual characteristics



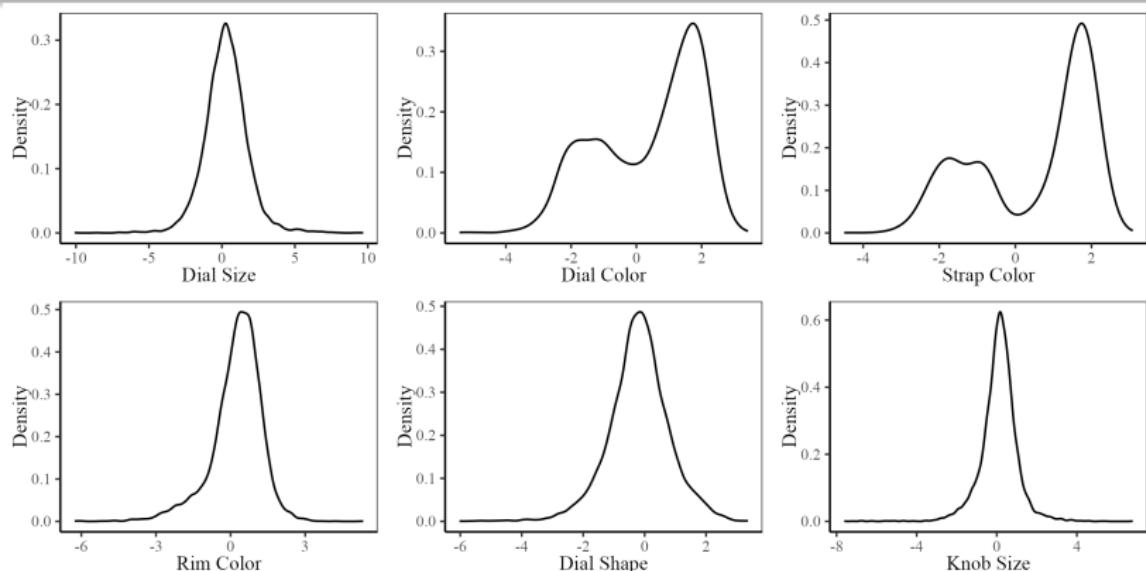
# Discovered Visual characteristics



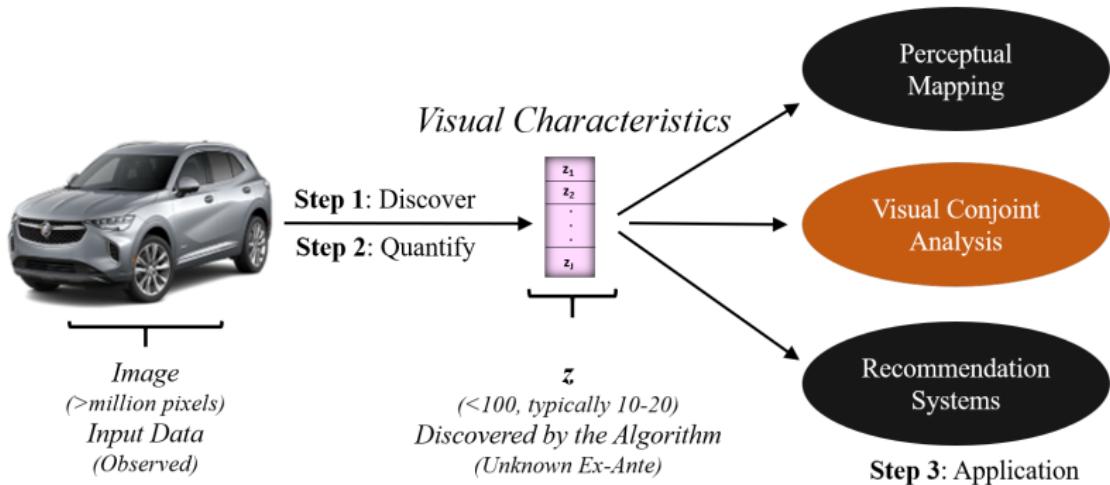
# Density of Discovered Visual characteristics



# Density of Discovered Visual characteristics



# Research Goals



- Visual conjoint has been challenging to do. With disentanglement we can create counterfactual designs to span the space.

# Conjoint Model Accuracy

## Generated Watches

Model	Out-of-Sample Hit Rate (SD)
Disentangled Embedding + Logit Model (-)	63.16% (2.34%)
Disentangled Embedding + Neural Net (-)	65.81% (2.22%)
Pretrained Deep Learning Model Embedding (O)	68.31% (1.54%)
Disentangled Embedding + Neural Net (O)	67.52% (0.92%)
Disentangled Embedding + Random Forest (O)	68.77% (0.90%)
Disentangled Embedding + XGBoost (O)	69.10% (0.41%)
<b>Disentangled Embedding + HB Model (O + U)</b>	<b>71.61% (1.87%)</b>
Disentangled Embedding + HB Model + Interactions (O + U)	70.68% (1.35%)

- Pretrained Deep learning model is trained on *millions of images*, and has millions of parameters
- Our Hierarchical Bayes (HB) model has a small number parameters, and all predictions are based on only 6 visual characteristics
- With 6 visual characteristics, our HB model outperforms the pretrained deep neural net

# Ideal Points for Segments

- Discover 2 segments with distinct and differentiated preferences

# Ideal Points for Segments

- Discover 2 segments with distinct and differentiated preferences



Segment 1:  
“Ideal Point” Watch Design



Segment 2:  
“Ideal Point” Watch Design

# Conclusion

We obtain interpretable visual characteristics directly from unstructured product images

- *automatically discover (extract) characteristics*

# Conclusion

We obtain interpretable visual characteristics directly from unstructured product images

- *automatically discover* (extract) characteristics
- *quantify* these characteristics

# Conclusion

We obtain interpretable visual characteristics directly from unstructured product images

- *automatically discover* (extract) characteristics
- *quantify* these characteristics
- *generate visual design that span the space of visual characteristics*

# Conclusion

We obtain interpretable visual characteristics directly from unstructured product images

- *automatically discover* (extract) characteristics
- *quantify* these characteristics
- *generate* visual design that span the space of visual characteristics

# Conclusion

We obtain interpretable visual characteristics directly from unstructured product images

- *automatically discover* (extract) characteristics
- *quantify* these characteristics
- *generate* visual design that span the space of visual characteristics

## Applications

We then used the model to:

- generate new counterfactual designs to obtain consumer preferences over visual characteristics.
- obtain ideal point visual designs corresponding to different consumer segments

# Market Structure Mapping with Interpretable Visual Characteristics

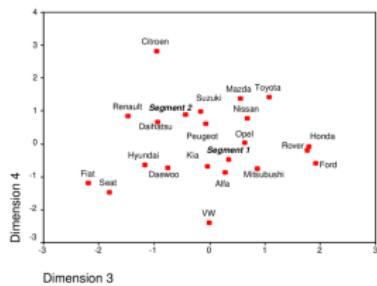
Sisodia and Kumar

Presenting at:  
*Singapore Management University*  
*Lee Kong Chian School of Business*  
September 2025

# What is a Market Structure Map?

A common element of market structure analysis is the derivation of a **market structure map**, that is, a spatial representation of firms' competitive positions relative to one another based on some measure of their competitive relationships (DeSarbo et al. 1993)

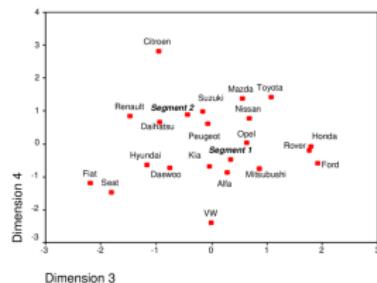
- Graphical representations of product positioning (e.g. Multidimensional Scaling)



# What is a Market Structure Map?

A common element of market structure analysis is the derivation of a **market structure map**, that is, a spatial representation of firms' competitive positions relative to one another based on some measure of their competitive relationships (DeSarbo et al. 1993)

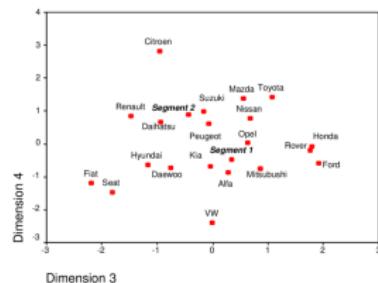
- Graphical representations of product positioning (e.g. Multidimensional Scaling)
- Strategic tool for understanding competitive landscape that enable us to:



# What is a Market Structure Map?

A common element of market structure analysis is the derivation of a **market structure map**, that is, a spatial representation of firms' competitive positions relative to one another based on some measure of their competitive relationships (DeSarbo et al. 1993)

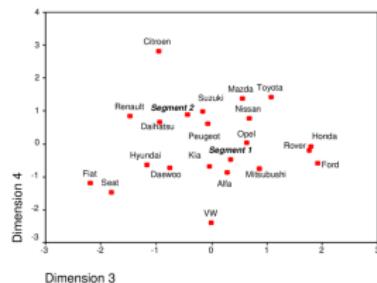
- Graphical representations of product positioning (e.g. Multidimensional Scaling)
- Strategic tool for understanding competitive landscape that enable us to:
  - Identify market gaps



# What is a Market Structure Map?

A common element of market structure analysis is the derivation of a **market structure map**, that is, a spatial representation of firms' competitive positions relative to one another based on some measure of their competitive relationships (DeSarbo et al. 1993)

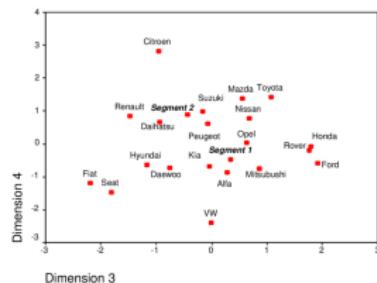
- Graphical representations of product positioning (e.g. Multidimensional Scaling)
- Strategic tool for understanding competitive landscape that enable us to:
  - Identify market gaps
  - Spot new opportunities



# What is a Market Structure Map?

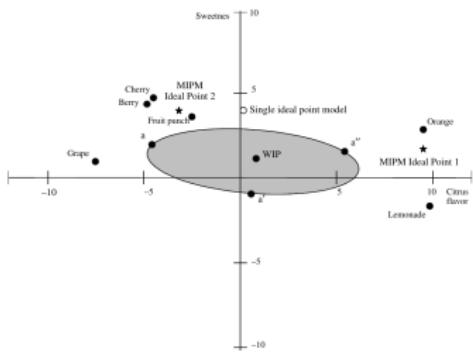
A common element of market structure analysis is the derivation of a **market structure map**, that is, a spatial representation of firms' competitive positions relative to one another based on some measure of their competitive relationships (DeSarbo et al. 1993)

- Graphical representations of product positioning (e.g. Multidimensional Scaling)
- Strategic tool for understanding competitive landscape that enable us to:
  - Identify market gaps
  - Spot new opportunities
  - **Identify competitors and evaluate positioning**

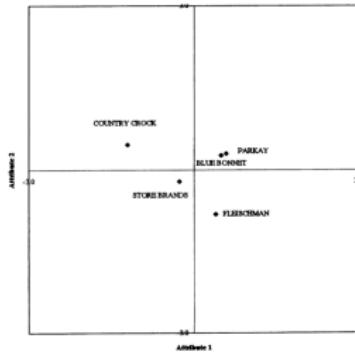


# Market Structure Map: Examples

Compositional (Left) & Decompositional (Right)



Lee, Sudhir and Steckel (2002), *Journal of Marketing Research*



Erdem (1996), *Marketing Science*

# Market Structure Mapping

## Sources of Data

### Compositional (Building up)

- Product characteristics – functional, psychological

### Decompositional (Breaking down)

- Survey data by asking 1st and 2nd choices

# Market Structure Mapping

## Sources of Data

### Compositional (Building up)

- Product characteristics – functional, psychological

### Decompositional (Breaking down)

- Survey data by asking 1st and 2nd choices
- Sales data (cross-elasticity of demand)

# Market Structure Mapping

## Sources of Data

### Compositional (Building up)

- Product characteristics – functional, psychological

### Decompositional (Breaking down)

- Survey data by asking 1st and 2nd choices
- Sales data (cross-elasticity of demand)
- **User comments and reviews**

# Market Structure Mapping

## Sources of Data

### Compositional (Building up)

- Product characteristics – functional, psychological

### Decompositional (Breaking down)

- Survey data by asking 1st and 2nd choices
- Sales data (cross-elasticity of demand)
- User comments and reviews
- Search Co-occurrence

# Market Structure Mapping

## Sources of Data

### Compositional (Building up)

- Product characteristics – functional, psychological

### Decompositional (Breaking down)

- Survey data by asking 1st and 2nd choices
- Sales data (cross-elasticity of demand)
- User comments and reviews
- Search Co-occurrence

# Market Structure Mapping

## Sources of Data

### Compositional (Building up)

- Product characteristics – functional, psychological

### Decompositional (Breaking down)

- Survey data by asking 1st and 2nd choices
- Sales data (cross-elasticity of demand)
- User comments and reviews
- Search Co-occurrence

### What's common?

Neither of these approaches have considered visual characteristics.



# Are these cars close substitutes?

First, consider *functional* product characteristics:

(A)

\$179,000-\$243,000, 553 HP,  
16-23 MPG

(B)

From \$242,700, 542-650 HP,  
14-22 MPG

# Are these cars close substitutes?

First, consider *functional* product characteristics:

(A)

\$179,000-\$243,000, 553 HP,  
16-23 MPG

Ferrari California



(B)

From \$242,700, 542-650 HP,  
14-22 MPG

Bentley Continental GT



## Importance of Visual Design

Functional characteristics alone are *not sufficient* to even place the car in an appropriate segment (submarket)

# Visual design matters \*a lot\* for cars



*“Exterior look/design is the top reason shoppers avoid a particular vehicle (30%), followed by cost (17%).”*

*—JD Power Avoider Study 2015*

# Challenges in using visual images for market structure mapping

## Three Challenges:

- ① Inputs from Raw Images: Backgrounds, lighting, and camera angles overshadow actual product design, causing mis-clusters.

# Challenges in using visual images for market structure mapping

## Three Challenges:

- ① Inputs from Raw Images: Backgrounds, lighting, and camera angles overshadow actual product design, causing mis-clusters.
- ② Scalability: Need to work with hundreds or thousands of product models / images, so challenging to have human intervention

# Challenges in using visual images for market structure mapping

## Three Challenges:

- ① Inputs from Raw Images: Backgrounds, lighting, and camera angles overshadow actual product design, causing mis-clusters.
- ② Scalability: Need to work with hundreds or thousands of product models / images, so challenging to have human intervention
- ③ Black-Box (Un)interpretable Embeddings: Pretrained networks may focus on brand logos or color, complicating interpretation of true style.

# Challenges in using visual images for market structure mapping

## Three Challenges:

- ① Inputs from Raw Images: Backgrounds, lighting, and camera angles overshadow actual product design, causing mis-clusters.
- ② Scalability: Need to work with hundreds or thousands of product models / images, so challenging to have human intervention
- ③ Black-Box (Un)interpretable Embeddings: Pretrained networks may focus on brand logos or color, complicating interpretation of true style.
  - We need to be able to choose what visual characteristics matter for market structure

# Challenges in using visual images for market structure mapping

## Three Challenges:

- ① Inputs from Raw Images: Backgrounds, lighting, and camera angles overshadow actual product design, causing mis-clusters.
- ② Scalability: Need to work with hundreds or thousands of product models / images, so challenging to have human intervention
- ③ Black-Box (Un)interpretable Embeddings: Pretrained networks may focus on brand logos or color, complicating interpretation of true style.
  - We need to be able to choose what visual characteristics matter for market structure
  - Does color matter? Does angle matter?

# Should Color be included in Visual Characteristics Maps?

- Should a red Ferrari California be close substitute to a red Toyota Corolla?

- **UK Automobile Market**
  - 2,439 make-model-year observations from 2008 to 2017
  - 379 unique models from 49 unique makes
- **Sales and Auto Characteristics:**
  - Characteristics (price, MPG, horsepower, weight, dimensions, etc.) from Parker's
  - Sales from DVM-CAR<sup>2</sup>
- **Images:**
  - Front-facing automobile images from DVM-CAR
  - Converted from color to grayscale to *focus on shape* rather than color.
  - (very) Low-resolution:  $128 \times 128$  pixels

---

<sup>2</sup>

Huang, Jingmin, et al. (2022), *IEEE International Conference on Big Data*

# Segments in the UK automobile market

Segment	Description				
A	Minicars		Seg A (Minicar)		
B	Subcompact		Seg B (Subcompact)		
C	Compact		Seg C (Compact)		
D	Mid-size		Seg D (Mid-Sized)		
E	Mid-size Luxury		Seg E (Mid-Sized Luxury)		
J	SUV		Seg J (SUV)		
M	MPV		Seg M (MPV)		

# Summary Statistics of the 2013 UK Auto Market

Variable	Mean	St. Dev.	Min.	Max.
Total Quantity Sold	8,074.834	13,714.100	1	113,390
Price (in £000 units)	26.333	14.668	7.868	108.624
MPG (tens of miles per gallon)	4.999	1.058	2.250	7.200
Weight (in 10 lbs)	327.704	2.444	324.506	332.106
HP/Wt (in HP per 10 lbs.)	0.461	0.169	0.060	1.347
Length (in 1000 inches)	1.724	0.178	1.062	2.054
Width (in 1000 inches)	0.756	0.063	0.580	0.899
Height (in 1000 inches)	0.616	0.054	0.537	0.780
Wheelbase (in 1000 inches)	1.046	0.081	0.735	1.266
Space (length × width)	1.310	0.215	0.697	1.759
Vehicle Segment (Proportion):				
Segment A (Minicars)	0.112	0.316	0	1
Segment B (Subcompact)	0.120	0.326	0	1
Segment C (Compact)	0.162	0.369	0	1
Segment D (Mid-Size)	0.129	0.335	0	1
Segment E (Mid-Size Luxury)	0.071	0.257	0	1
Segment J (SUV)	0.241	0.428	0	1
Segment M (MPV)	0.166	0.373	0	1
Country of Origin (Proportion):				
France	0.108	0.311	0	1
Germany	0.241	0.428	0	1
Japan	0.220	0.415	0	1
South Korea	0.091	0.289	0	1
UK	0.108	0.311	0	1
USA	0.058	0.234	0	1

# Match each model with front-facing image

Is this sufficient?

*"Market research studies have shown that 70% of consumers identify and judge automobiles by the appearance of headlights and grille located on the face of the automobile."*

– *The Wall Street Journal*  
2006



# Loss Function & Supervisory Signals

$$\underbrace{L(\theta, \phi; \mathbf{m}, \mathbf{v})}_{\text{Disentanglement Loss}} = \underbrace{-\mathbb{E}_{q_\phi(\mathbf{v}|\mathbf{m})} [\log p_\theta(\mathbf{m}|\mathbf{v})]}_{\text{Reconstruction Loss}} + \underbrace{I_q(\mathbf{v}, \mathbf{m})}_{\text{Mutual Information Loss}} \\
 + \lambda_1 \underbrace{KL \left[ q(\mathbf{v}) || \prod_{j=1}^J q(v_j) \right]}_{\text{Total Correlation Loss}} + \underbrace{\sum_{j=1}^J KL [q(v_j) || p(v_j)]}_{\text{Dimension-Wise KL Divergence Loss}} + \lambda_2 \underbrace{P(\widehat{\mathbf{y}(\mathbf{v})}, \mathbf{y})}_{\text{Supervised Loss}}$$
(1)

Number of Signals	Supervisory Signals	$\lambda_1$	$\lambda_2$	UDR
3	Wheelbase, Width, Height	50	10	0.739
3	HP/Weight, MPG, Space	50	30	0.710
1	Price	50	30	0.708
1	Weight	50	40	0.708
1	Wheelbase	50	30	0.690
1	Width	50	5	0.689
3	Length, Width, Height	50	40	0.678
1	Length	50	40	0.666
0	Unsupervised $\beta$ -TCVAE	50	0	0.658
1	Height	30	20	0.378
1	Country of Origin	10	10	0.139
1	Segment	10	10	0.134
1	Unsupervised VAE	1	0	0.073
1	Unsupervised AE	0	0	0.074
1	Make	1	1	0.072

$$\lambda_1 \in [1, 5, 10, 20, 30, 40, 50] \text{ and } \lambda_2 \in [0, 1, 5, 10, 20, 30, 40, 50].$$

# Visual Characteristic 1: Body Shape



Correlation Between Discovered Visual Dimensions and Physical Vehicle Measures

	Wheelbase	Weight	Length	Height	Width	Height/Width Ratio
Body Shape	0.30	0.33	0.39	-0.28	0.25	-0.42

- Hatchback-like profiles → smaller wheelbase, lower length, and lighter
- Hatchback-like profiles → taller and narrower

# Visual Characteristic 2: Boxiness



Correlation Between Discovered Visual Dimensions and Physical Vehicle Measures

	Wheelbase	Weight	Length	Height	Width	Height/Width Ratio
Boxiness	0.05	-0.07	0.14	-0.59	0.02	-0.49

- Higher degree of boxiness → Taller and upfront from the front
- Lower degree of boxiness → Flatter and sleeker
- Boxiness captures cabin “uprightness” (not related to length or wheelbase)

# Visual Characteristic 3: Grille Height



Correlation Between Discovered Visual Dimensions and Physical Vehicle Measures

	Wheelbase	Weight	Length	Height	Width	Height/Width Ratio
Grille Height	0.04	0.02	0.05	-0.04	0.03	-0.05

- Reflects stylistic design choices
- Unrelated to size-related factors

# Visual Characteristic 4: Grille Width



Correlation Between Discovered Visual Dimensions and Physical Vehicle Measures

	Wheelbase	Weight	Length	Height	Width	Height/Width Ratio
Grille Width	0.12	0.08	0.12	0.03	0.15	-0.09

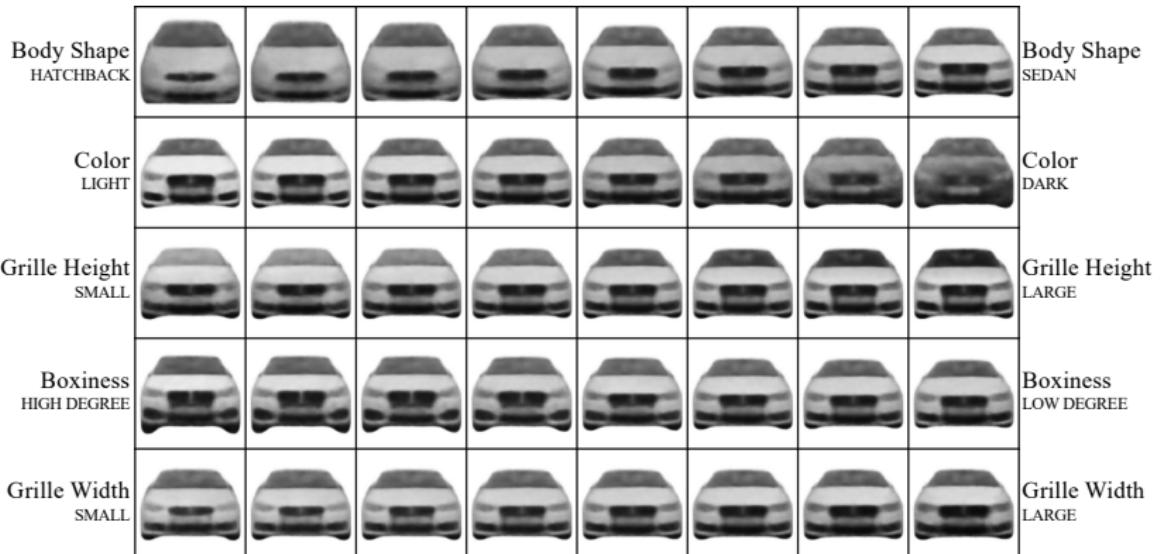
- Reflects stylistic design choices
- Unrelated to size-related factors

# Visual Characteristic 5: Color



- We find the visual characteristic of *Color*.
- Color should not impact market structure because a car is available in many colors.
- We, therefore, exclude it when we create market structure maps.
- This is only possible because of disentanglement.

# Disentanglement obtains four interpretable visual characteristics



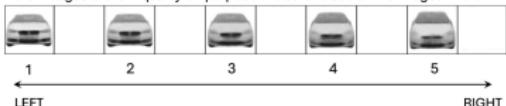
# Human Understanding of Visual Characteristics

- Showed respondents sequence of car images

Q1/4: Look at the below image to see the various parts of a car.



Now, carefully examine each car image below from 1 to 5, going from left to right.  
Note: Images are low-quality on purpose. Be sure to see all the images 1 to 5.



How does the car change the most as you go from image 1 to 5? Go through each part of the car one by one before deciding your response. Write it in a few words.

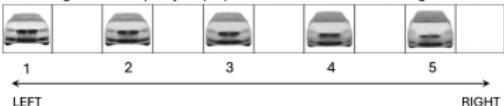
# Human Understanding of Visual Characteristics

- Showed respondents sequence of car images
- Asked respondents open-ended questions about visual changes

Q1/4: Look at the below image to see the various parts of a car.



Now, carefully examine each car image below from 1 to 5, going from left to right.  
Note: Images are low-quality on purpose. Be sure to see all the images 1 to 5.



How does the car change the most as you go from image 1 to 5? Go through each part of the car one by one before deciding your response. Write it in a few words.

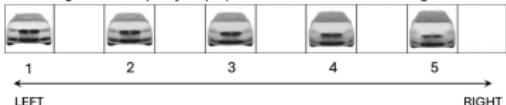
# Human Understanding of Visual Characteristics

- Showed respondents sequence of car images
- Asked respondents open-ended questions about visual changes
- Used LLMs (GPT-4, Claude) to identify *common themes* in respondent responses

Q1/4: Look at the below image to see the various parts of a car.



Now, carefully examine each car image below from 1 to 5, going from left to right.  
Note: Images are low-quality on purpose. Be sure to see all the images 1 to 5.



How does the car change the most as you go from image 1 to 5? Go through each part of the car one by one before deciding your response. Write it in a few words.

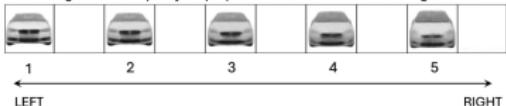
# Human Understanding of Visual Characteristics

- Showed respondents sequence of car images
- Asked respondents open-ended questions about visual changes
- Used LLMs (GPT-4, Claude) to identify *common themes* in respondent responses
- Example: LLM summarizes that respondents are saying that cars becoming lower, flatter, and wider can be thought of as boxiness

Q1/4: Look at the below image to see the various parts of a car.



Now, carefully examine each car image below from 1 to 5, going from left to right.  
Note: Images are low-quality on purpose. Be sure to see all the images 1 to 5.



How does the car change the most as you go from image 1 to 5? Go through each part of the car one by one before deciding your response. Write it in a few words.

# Quantification Validation Survey Question

Which pair of cars in your judgment are visually more similar? Carefully check both large and small visual aspects. Do not consider any non-visual features like brand or price.



Left Pair



Right Pair



# Correlation Between Different Characteristics

	Price	Functional Characteristics			Boxiness	Visual Characteristics		Grille Width
		MPG	HP/Weight	Space		Body Shape	Grille Height	
Price	1.00							
MPG	-0.60	1.00						
HP/Weight	0.74	-0.48	1.00					
Space	0.67	-0.47	0.36	1.00				
Boxiness	0.06	0.04	0.29	0.09	1.00			
Body Shape	0.50	-0.25	0.54	0.36	0.13	1.00		
Grille Height	0.11	0.03	0.12	0.05	0.04	-0.02	1.00	
Grille Width	0.07	-0.05	0.04	0.15	0.01	-0.12	-0.05	1.00

# Does Form Follow Function?

Insight 1: Products close in functional space are differentiated visually

- *Operationalization:* Calculate correlation between distances in functional product characteristics and visual product characteristics for each pair of make-models within each segment

# Does Form Follow Function?

Insight 1: Products close in functional space are differentiated visually

- *Operationalization:* Calculate correlation between distances in functional product characteristics and visual product characteristics for each pair of make-models within each segment
- *Result:* Low correlation between functional and visual similarity

# Does Form Follow Function?

Insight 1: Products close in functional space are differentiated visually

- *Operationalization:* Calculate correlation between distances in functional product characteristics and visual product characteristics for each pair of make-models within each segment
- *Result:* Low correlation between functional and visual similarity
- *Implication:* Visual can create distinctiveness

# Does Form Follow Function?

Insight 1: Products close in functional space are differentiated visually

- *Operationalization:* Calculate correlation between distances in functional product characteristics and visual product characteristics for each pair of make-models within each segment
- *Result:* Low correlation between functional and visual similarity
- *Implication:* Visual can create distinctiveness

# Does Form Follow Function?

## Insight 1: Products close in functional space are differentiated visually

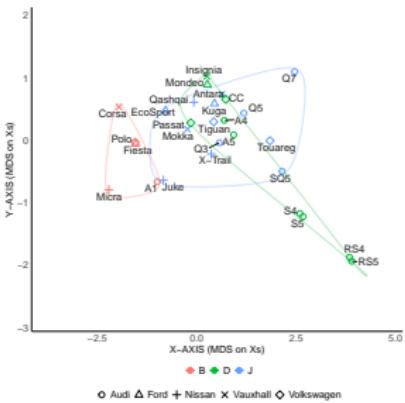
- *Operationalization:* Calculate correlation between distances in functional product characteristics and visual product characteristics for each pair of make-models within each segment
- *Result:* Low correlation between functional and visual similarity
- *Implication:* Visual can create distinctiveness

Segment	Correlation	Std. Error
A (Minicars)	-0.08	0.05
B (Subcompact)	-0.05	0.05
C (Compact)	-0.05	0.04
D (Mid-size)	0.02	0.04
E (Mid-size Luxury)	0.13	0.10
J (SUV)	0.09	0.03
M (MPV)	0.11	0.04

# Market Structure Maps

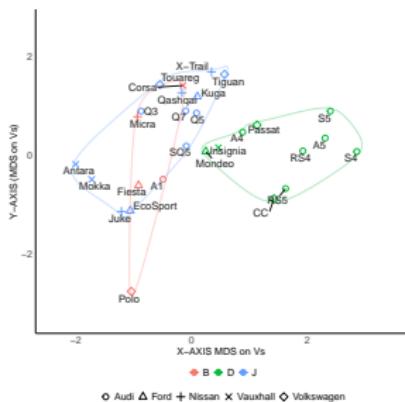
# Market Structure Map using only functional characteristics

Segment	Description
B	Subcompact
D	Mid-size
J	SUV



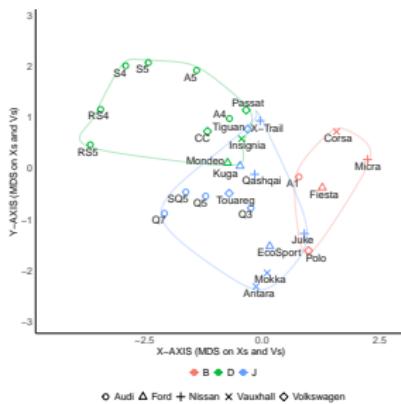
# Market Structure Map using only visual characteristics

Segment	Description
B	Subcompact
D	Mid-size
J	SUV



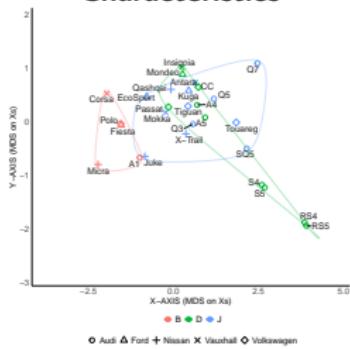
# Market Structure Map using both type of characteristics

Segment	Description
B	Subcompact
D	Mid-size
J	SUV

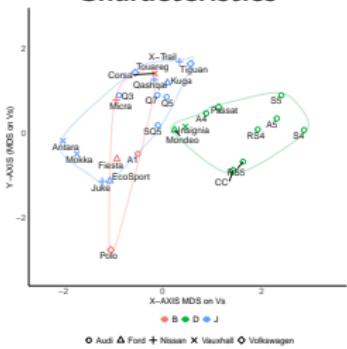


# Insight 2: Visual increases differentiation and helps separate segments

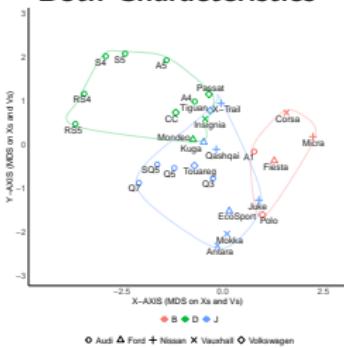
Only Functional  
Characteristics



Only Visual  
Characteristics



Both Characteristics

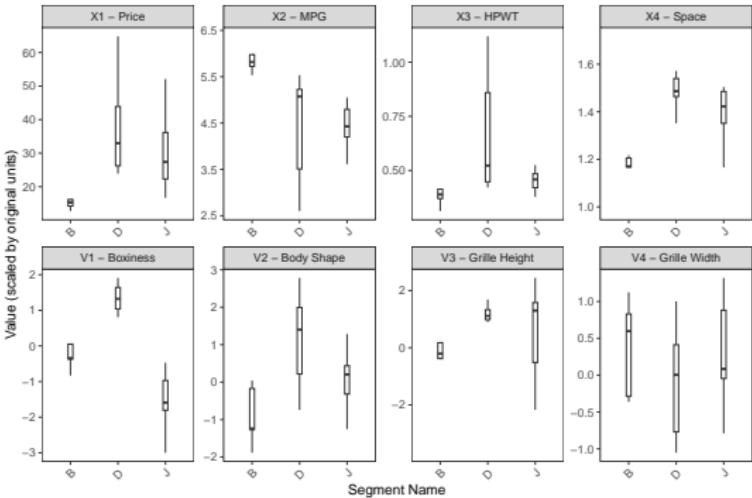


8.3% segment overlap

# Does Disentanglement Provide Further Insights?

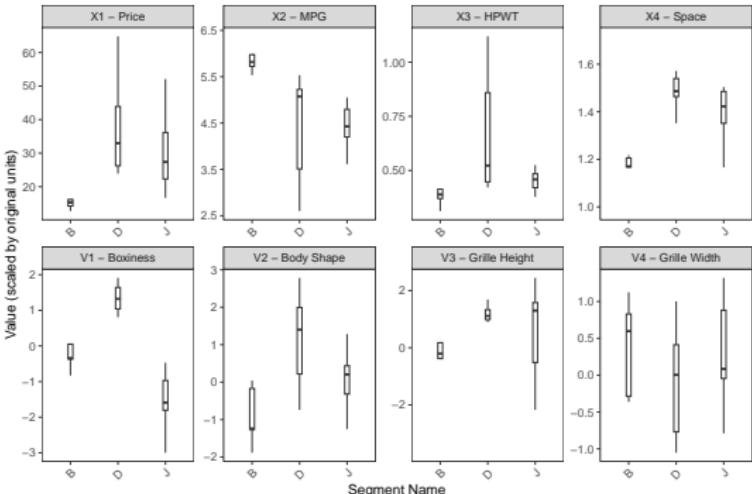
# Insight 2D: Which visual characteristics increase differentiation?

- Disentanglement identifies interpretable characteristics contributing to differentiation



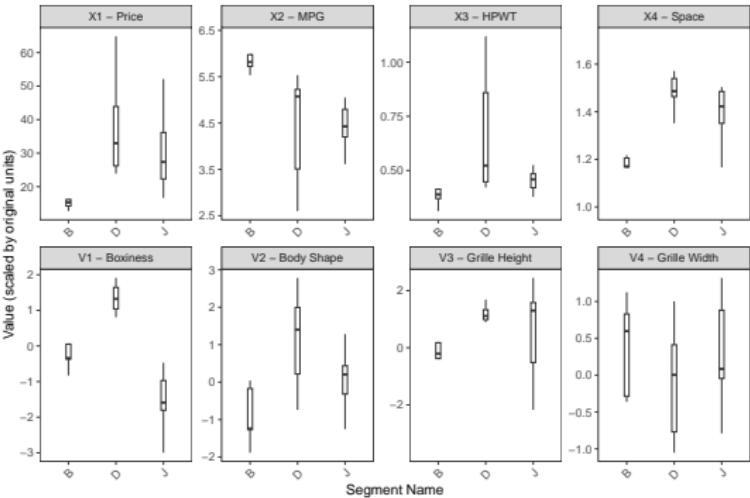
# Insight 2D: Which visual characteristics increase differentiation?

- Disentanglement identifies interpretable characteristics contributing to differentiation
- Seg D & J overlap on most of the functional characteristics but Seg B is different



# Insight 2D: Which visual characteristics increase differentiation?

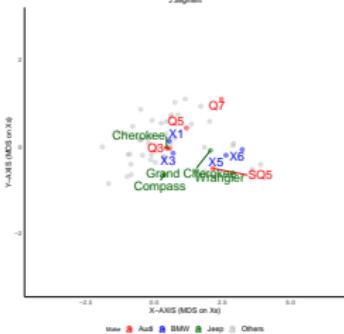
- Disentanglement identifies interpretable characteristics contributing to differentiation
- Seg D & J overlap on most of the functional characteristics but Seg B is different
- Seg D & J are different in boxiness and grille



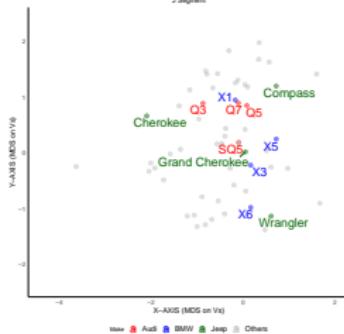
# Insight 3: Different Visual Strategy Across Brands

## Segment J – SUVs

**Only Functional  
Characteristics**

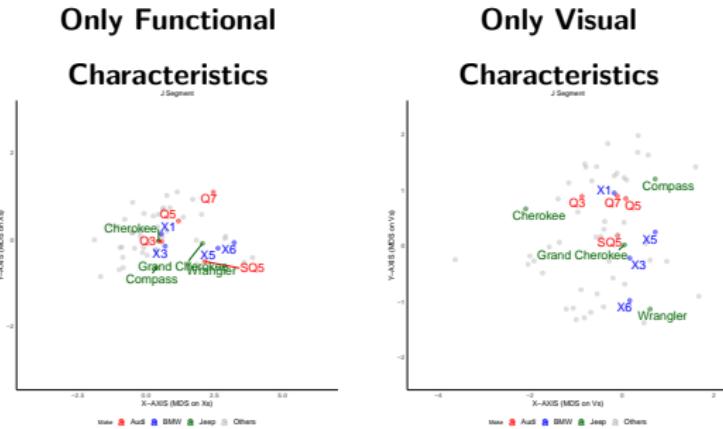


**Only Visual  
Characteristics**



# Insight 3: Different Visual Strategy Across Brands

## Segment J – SUVs



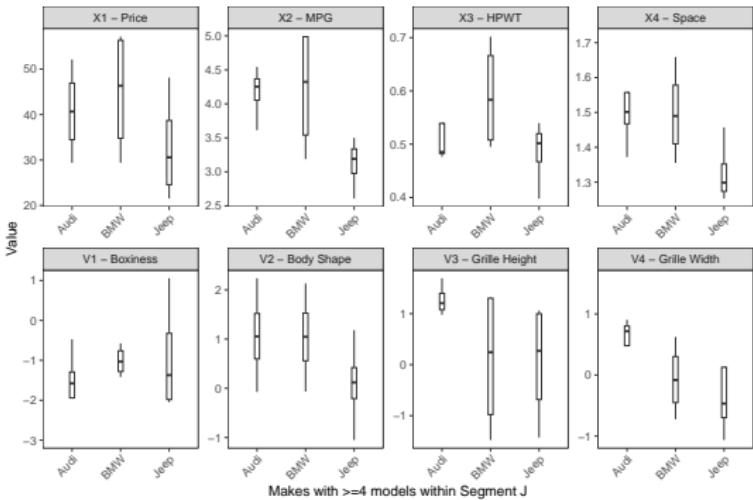
Area Share of a Make in Functional Space & Visual Space (Segment J: SUVs)

Make	Models	Area Share (Functional)	Area Share (Visual)	Ratio
Audi	4	17.90%	2.96%	6.05
BMW	4	6.35%	6.48%	0.98
Jeep	4	9.38%	28.42%	0.33

# Insight 3D: Different Visual Strategy Across Brands

Disentanglement identifies interpretable characteristics

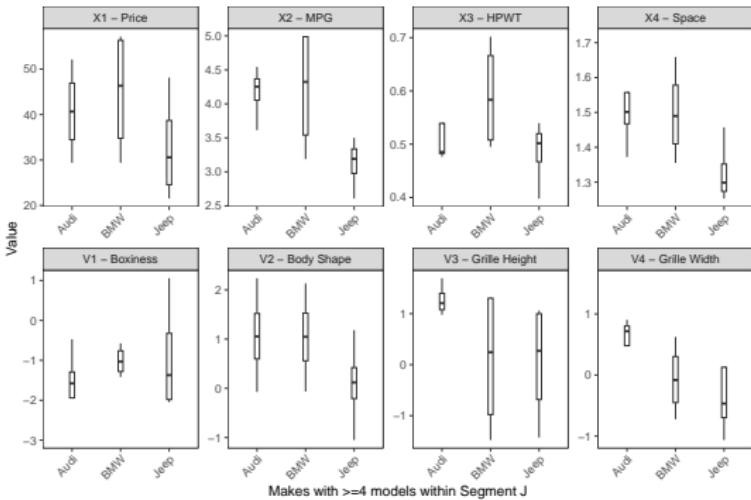
- Audi is very tight on some visual characteristics, notably, grille height and grille width  $\implies$  distinctive look



# Insight 3D: Different Visual Strategy Across Brands

Disentanglement identifies interpretable characteristics

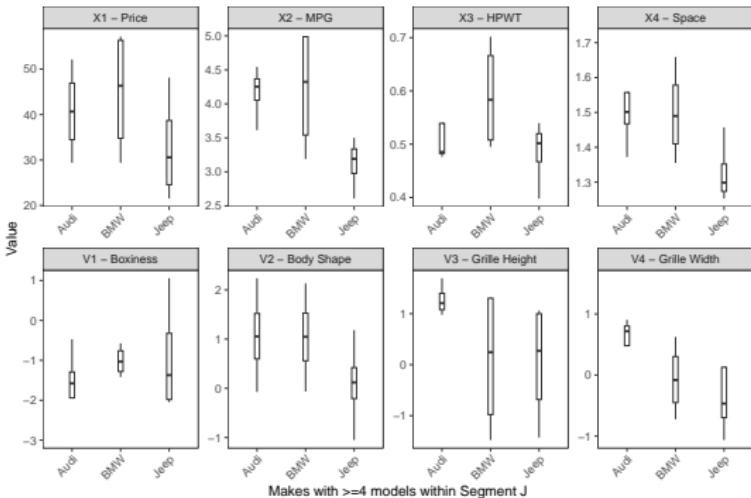
- Audi is very tight on some visual characteristics, notably, grille height and grille width  $\implies$  *distinctive look*
- BMW is typically a more “boxy” look



# Insight 3D: Different Visual Strategy Across Brands

Disentanglement identifies interpretable characteristics

- Audi is very tight on some visual characteristics, notably, grille height and grille width  $\implies$  *distinctive look*
- BMW is typically a more “boxy” look
- Jeep varies on most visual characteristics



# Insight 3: Different Visual Strategy Across Brands

Audi Q3



Audi Q5



Audi Q7



Audi SQ5



BMW X1



BMW X3



BMW X5



BMW X6



Jeep Cherokee



Jeep Compass



Jeep Grand  
Cherokee

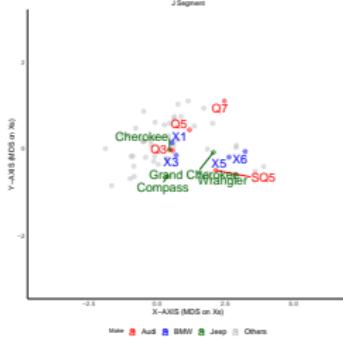


Jeep Wrangler

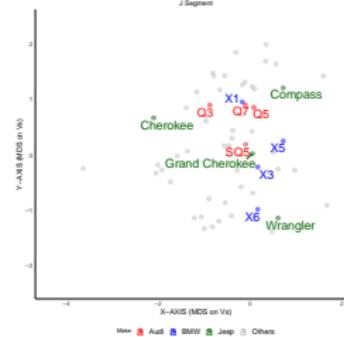


# Insight 4: Product-Level

## Only Functional Characteristics



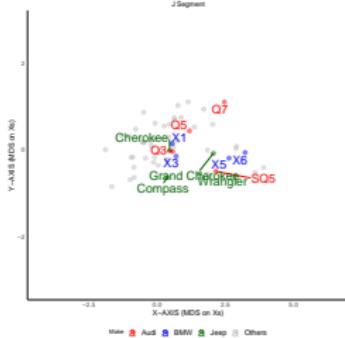
## Only Visual Characteristics



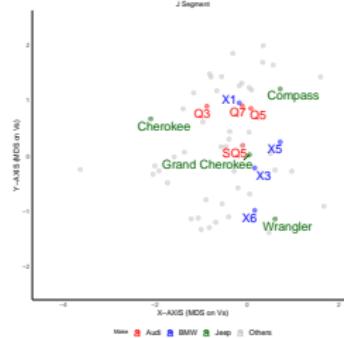
- Looking only at functional characteristics, BMW seems to be positioning X5 and X6 too close (cannibalization risk)

# Insight 4: Product-Level

## Only Functional Characteristics



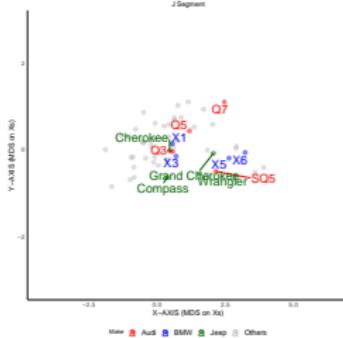
## Only Visual Characteristics



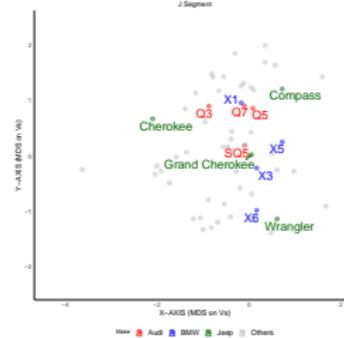
- Looking only at functional characteristics, BMW seems to be positioning X5 and X6 too close (cannibalization risk)
  - Looking at visual, we can see that they are now separated more

# Insight 4: Product-Level

Only Functional  
Characteristics



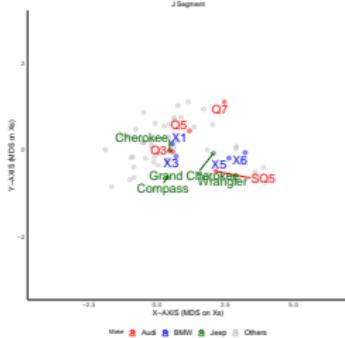
Only Visual  
Characteristics



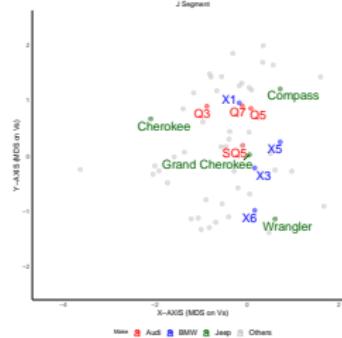
- Looking only at functional characteristics, BMW seems to be positioning X5 and X6 too close (cannibalization risk)
  - Looking at visual, we can see that they are now separated more
- Positioning BMW X3: Should I focus on Cherokee (which is what functional map suggests)?

# Insight 4: Product-Level

Only Functional  
Characteristics



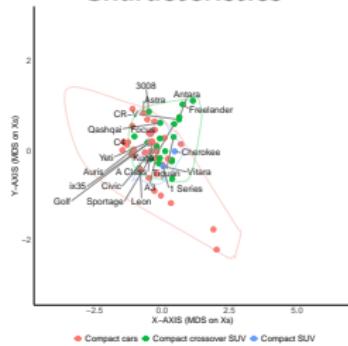
Only Visual  
Characteristics



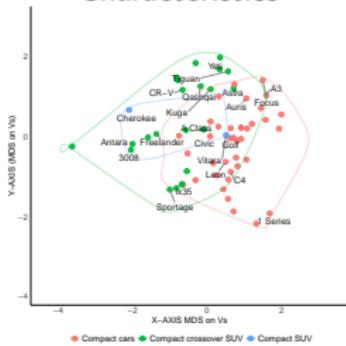
- Looking only at functional characteristics, BMW seems to be positioning X5 and X6 too close (cannibalization risk)
  - Looking at visual, we can see that they are now separated more
- Positioning BMW X3: Should I focus on Cherokee (which is what functional map suggests)?
  - Visual map suggests the Grand Cherokee is closer instead

# Insight 5: Leaping to Another Category

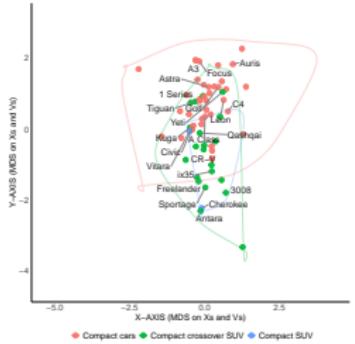
**Only Functional  
Characteristics**



**Only Visual  
Characteristics**

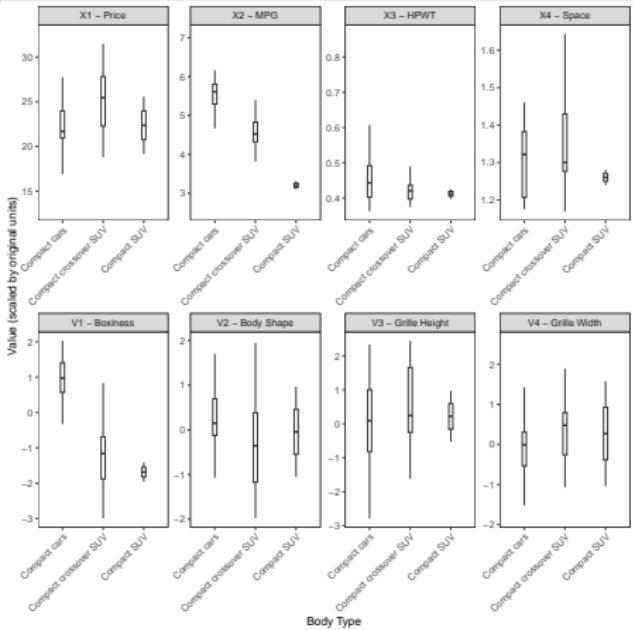


**Both Characteristics**



# Insight 5D: Leaping to Another Category

- Compact cars and compact crossover SUVs differ in boxiness
- Compact cars and compact crossover SUVs overlap across most functional characteristics
- Compact crossover SUVs and compact SUVs are similar in visual characteristics
- Compact crossover SUVs and compact SUVs differ in some



# Connecting Consumer Search to Market Structure Maps

# Consumer Search $\iff$ Market structure maps?

We use Google Trends to look for consumers searching pairs of models

- Consumer searching for “Honda Civic” and “Toyota Corolla” more likely to be comparing

$$\log_{10}(Y_{AB} + 1) = \beta_0 + \beta_1(|X_A - X_B|) + \beta_2(|V_A - V_B|) + \epsilon_{AB}$$

---

<sup>3</sup>Raw Google Trend values are scaled by each pair's segment sales share

# Consumer Search $\iff$ Market structure maps?

We use Google Trends to look for consumers searching pairs of models

- Consumer searching for “Honda Civic” and “Toyota Corolla” more likely to be comparing
- Models likely in consideration set

$$\log_{10}(Y_{AB} + 1) = \beta_0 + \beta_1(|X_A - X_B|) + \beta_2(|V_A - V_B|) + \epsilon_{AB}$$

---

<sup>3</sup>Raw Google Trend values are scaled by each pair's segment sales share

# Consumer Search $\iff$ Market structure maps?

We use Google Trends to look for consumers searching pairs of models

- Consumer searching for “Honda Civic” and “Toyota Corolla” more likely to be comparing
- Models likely in consideration set

$$\log_{10}(Y_{AB} + 1) = \beta_0 + \beta_1(|X_A - X_B|) + \beta_2(|V_A - V_B|) + \epsilon_{AB}$$

- $\log_{10}(Y_{AB} + 1)$  is the log-transformed normalized Google Trends score for the pair (model A, model B) belonging to the same segment<sup>3</sup>

---

<sup>3</sup>Raw Google Trend values are scaled by each pair's segment sales share

# Consumer Search $\iff$ Market structure maps?

We use Google Trends to look for consumers searching pairs of models

- Consumer searching for “Honda Civic” and “Toyota Corolla” more likely to be comparing
- Models likely in consideration set

$$\log_{10}(Y_{AB} + 1) = \beta_0 + \beta_1(|X_A - X_B|) + \beta_2(|V_A - V_B|) + \epsilon_{AB}$$

- $\log_{10}(Y_{AB} + 1)$  is the log-transformed normalized Google Trends score for the pair (model A, model B) belonging to the same segment<sup>3</sup>
- $(X_A, X_B)$  are standardized functional characteristics

---

<sup>3</sup>Raw Google Trend values are scaled by each pair's segment sales share

# Consumer Search $\iff$ Market structure maps?

We use Google Trends to look for consumers searching pairs of models

- Consumer searching for “Honda Civic” and “Toyota Corolla” more likely to be comparing
- Models likely in consideration set

$$\log_{10}(Y_{AB} + 1) = \beta_0 + \beta_1(|X_A - X_B|) + \beta_2(|V_A - V_B|) + \epsilon_{AB}$$

- $\log_{10}(Y_{AB} + 1)$  is the log-transformed normalized Google Trends score for the pair (model A, model B) belonging to the same segment<sup>3</sup>
- $(X_A, X_B)$  are standardized functional characteristics
- $(V_A, V_B)$  are standardized visual characteristics

---

<sup>3</sup>Raw Google Trend values are scaled by each pair's segment sales share

# Consumer Search $\iff$ Market structure maps?

	(1)	(2)	(3)	(4)
Constant	2.778*** (0.205)	9.509*** (1.425)	2.974*** (0.291)	8.871*** (1.479)
$\Delta$ HP/Weight	-2.395 (1.691)	-3.761** (1.460)	-2.886* (1.638)	-4.021*** (1.453)
$\Delta$ Space	-1.685 (1.068)	1.200 (0.851)	-1.444 (1.031)	1.431* (0.853)
$\Delta$ MPG	-0.601*** (0.199)	-0.407** (0.163)	-0.599*** (0.196)	-0.432*** (0.165)
$\Delta$ Price	-0.214*** (0.037)	-0.188*** (0.031)	-0.203*** (0.036)	-0.194*** (0.032)
$\Delta$ Boxiness			-0.263 (0.161)	0.249 (0.154)
$\Delta$ Bodyshape			0.149 (0.108)	-0.361** (0.144)
$\Delta$ Grille Height			-0.284*** (0.072)	-0.007 (0.065)
$\Delta$ Grille Width			0.324** (0.138)	-0.001 (0.145)
Make Fixed Effects	No	Yes	No	Yes
Observations	306	306	306	306
R <sup>2</sup>	0.168	0.590	0.247	0.602
Adjusted R <sup>2</sup>	0.157	0.559	0.227	0.566

Note: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01

Consumers search for visually similar cars even after accounting for functional similarity

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps
  - Method allows us to select only *relevant* visual characteristics

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps
  - Method allows us to select only *relevant* visual characteristics
- (Interpretable) Visual characteristics can provide an entirely new set of insights at segment (submarket) level, brand level and product level

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps
  - Method allows us to select only *relevant* visual characteristics
- (Interpretable) Visual characteristics can provide an entirely new set of insights at segment (submarket) level, brand level and product level
- **Visual design increases differentiation across segments**

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps
  - Method allows us to select only *relevant* visual characteristics
- (Interpretable) Visual characteristics can provide an entirely new set of insights at segment (submarket) level, brand level and product level
- Visual design increases differentiation across segments
- **Different visual strategy across brands**

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps
  - Method allows us to select only *relevant* visual characteristics
- (Interpretable) Visual characteristics can provide an entirely new set of insights at segment (submarket) level, brand level and product level
- Visual design increases differentiation across segments
- Different visual strategy across brands
  - Brand consistency (e.g. Audi)

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps
  - Method allows us to select only *relevant* visual characteristics
- (Interpretable) Visual characteristics can provide an entirely new set of insights at segment (submarket) level, brand level and product level
- Visual design increases differentiation across segments
- Different visual strategy across brands
  - Brand consistency (e.g. Audi)
  - Within-brand differentiation (e.g. Jeep) in design

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps
  - Method allows us to select only *relevant* visual characteristics
- (Interpretable) Visual characteristics can provide an entirely new set of insights at segment (submarket) level, brand level and product level
- Visual design increases differentiation across segments
- Different visual strategy across brands
  - Brand consistency (e.g. Audi)
  - Within-brand differentiation (e.g. Jeep) in design
- Product close in functional space often differentiated visually and the other way around

# Conclusion and Takeaways

- We quantify human-interpretable visual design characteristics and obtain market structure maps
  - Method allows us to select only *relevant* visual characteristics
- (Interpretable) Visual characteristics can provide an entirely new set of insights at segment (submarket) level, brand level and product level
- Visual design increases differentiation across segments
- Different visual strategy across brands
  - Brand consistency (e.g. Audi)
  - Within-brand differentiation (e.g. Jeep) in design
- Product close in functional space often differentiated visually and the other way around
- Consumers search for visually similar cars even after accounting for functional similarity