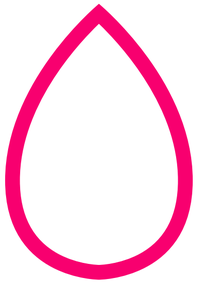
Edgar Anzaldua

User Research → UVP

Advancing Research 2021 | Day 2
13:50 - 14:25 PST
Version 3314

A story about research

- Using design thinking to do research
- Joining several research methods with a purpose in mind
- Focusing on stakeholders



Cars Guide™ .COM.AU

Canon

First...

How good are your **Personas**?

How is **Research** used?

Is the company **Confused**?

Are you a **Data Nerd**?

Based on a true story

The context

Doing
Research?

Using
Research?

Satisfied
with
research?

The context

UX activity is
design only

Product Managers
generalise users

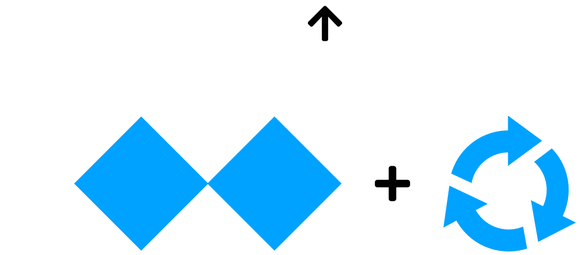
Maybe I can do **better** than that!

Research as a Product

Research as a Product

Research as a Product

Design Content Devs Marketing



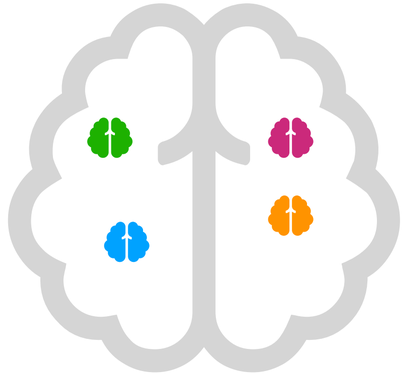
Research as a Product

Actionable

+

Stakeholder + User Centric

+



Shared Understanding

Principles

The right number of people

The right people

Easy to read, interpret & action

Roadmap

1. How can I get the right people?
2. How can I ask the right questions?
3. How can I segment them easily?
4. How can I consolidate findings to make it accesible, plain English, prioritisable and keep stakeholders engaged through the process?

How can I get the right people?

- Active,
- inactive or dormant,
- future, and
- former users.

Let's make a **survey**, and make **clusters** from the **data**!

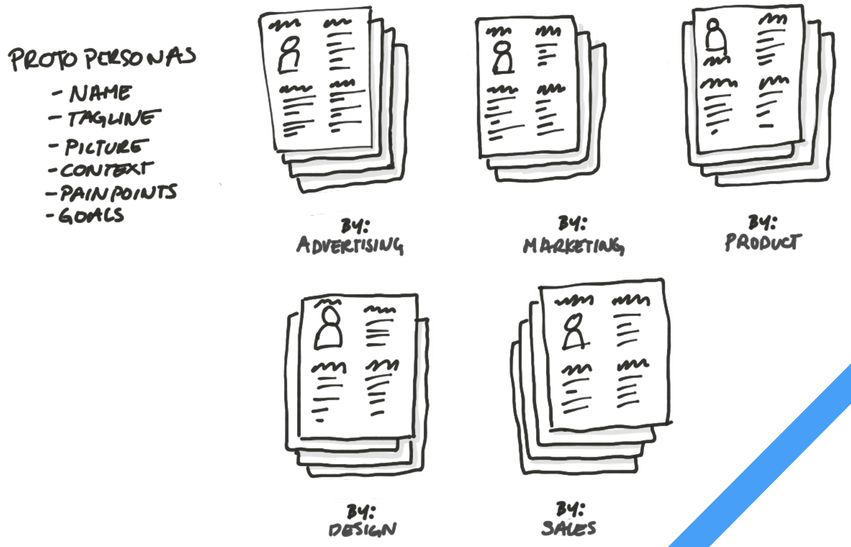
How can I ask the right questions?

Design Thinking to

- Ask the right questions and the questions in the right way
- Get good coverage
- Be minimal about it


Stakeholder Knowledge
1
Survey
2
Response Clusters
Screeners
3
Cluster Narratives
4
Personas + JM
UVP
Pain Points &
Goals
5
Business Model

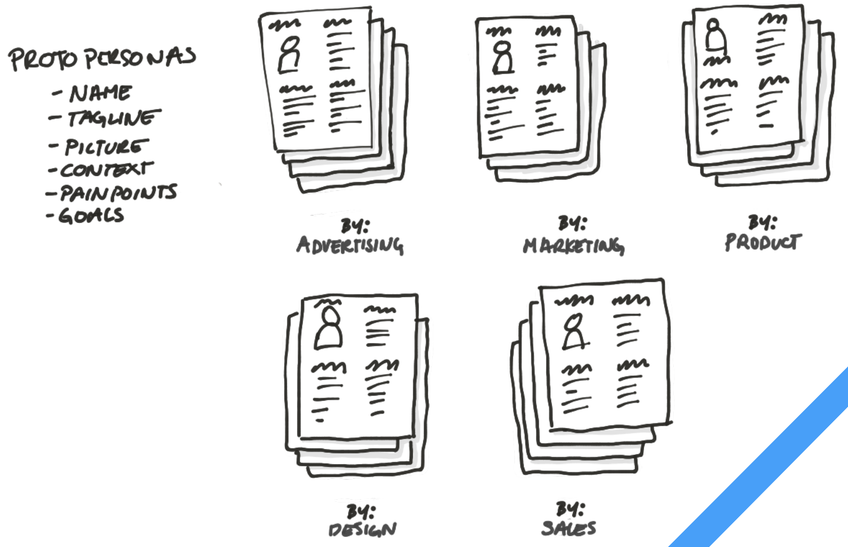
Remember?



Stakeholder Knowledge

5 Stakeholders
3 to 5 Proto-personas

Picture
Context
Pain-points
Goals
Name

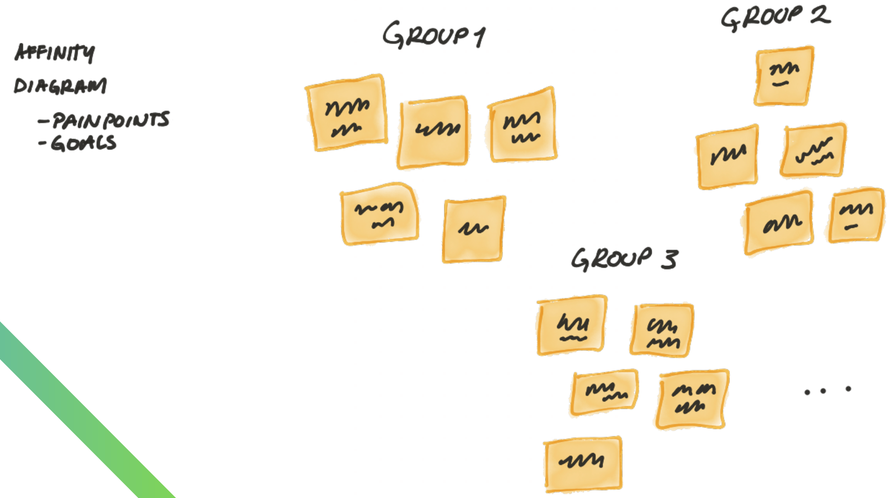


Stakeholder Knowledge

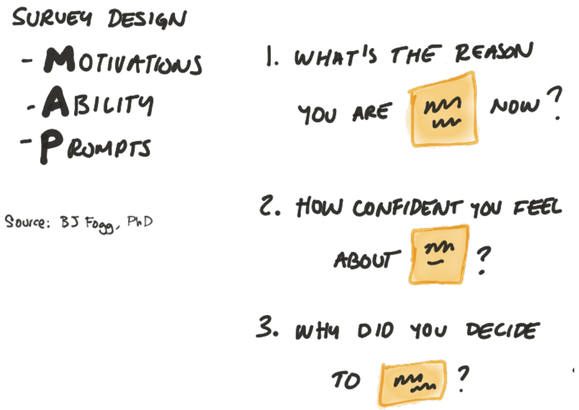
5 Stakeholders
3 to 5 Proto-personas
⌄
Picture
Context
Pain-points
Goals
Name

→

Affinity Mapping
⌄
26 Questions
2 Types of Questions



Survey



Response Requirements

500K

Unique monthly users

95%

Confidence level

5%

Margin of Error

384

Survey responses required

Source: https://www.surveysystem.com/sscalc.htm

That would be a lot of data to cluster manually!

Country Clustering Based on Search-Query Pattern Correlation

Edgar Alejandro Anzaldúa Moreno and Riaz Esmailzadeh, Carnegie Mellon University, Australia

Executive Summary

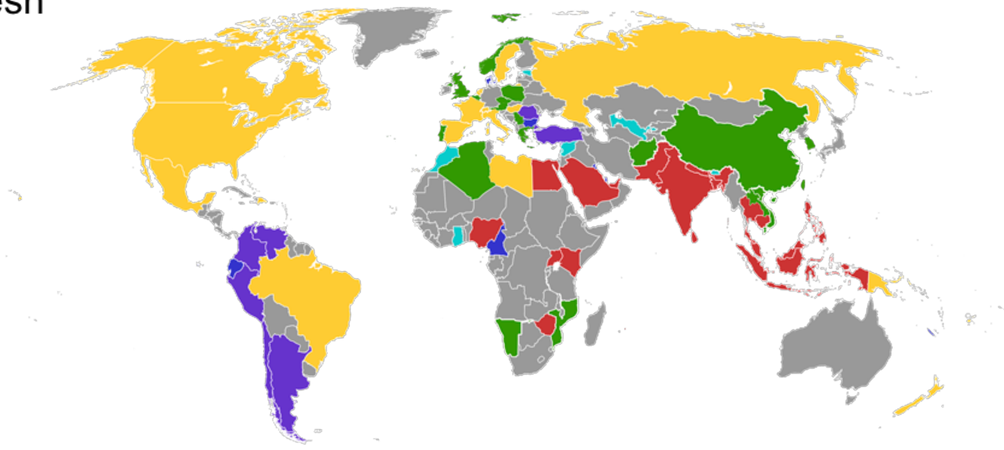
This paper introduces a technique to cluster regions or countries for marketing purposes. It uses information gathered through aggregated website search-query patterns, available from web logs or analytics software. Clustering done using this technique helps grouping countries that similarly respond to a particular marketing campaigns potentially reducing marketing costs. The method collects aggregated geographical and search-pattern information from keywords and the visit volume of each keyword. It uses abstract keyword logs in web analytics software to identify and understand target market segments more effectively. It is argued that actual search logs could be better indication of market behavior than language or cultural boundaries. By analyzing web logs with this technique, marketers will be able to plan marketing campaigns strategically based on the empirical data on website visitors. This data is derived from huge long-tailed keyword search sources, and ultimately helps understand and adapt better to customer's interests. The paper starts with an introduction on the underlying problem, the analysis method, and its four basic steps: preparing the data, creating a correlation matrix, identifying criteria and region clustering using the Louvain method of community structure over random networks. The results for a specific dataset are then presented and conclude with a list of future work steps. Results in this paper show that country-based keyword analysis can yield search-query patterns that unveil connections between culturally dissimilar regions.

Introduction

Thanks to the internet, people all over the world, access thousands of webpages every minute. There are several ways of accessing a website, ranging from direct access (like typing an URL on the internet browser's address bar) to using internet search engines that facilitate access to approximately 18.63 billion web pages[1] through manual and crawler-based indexing techniques. Technology, standards and software tools together allow tracing of this kind of user behavior by enabling website owners to record how people get into their website: whether s/he types the URL into the address bar or if s/he used a search engine such as Yahoo, Google, Bing or Ask to get to the website. If a search engine is used, the actual search term is of interest: i.e. what was the keyword[2] that brought that user to the website.

Most internet users are likely to use a search engine when trying to find a website for the first time. Even when people know the URL of a website, they sometimes write the name of the site, the products or the services that the website offers in the search field of a given search engine to access them. Organizing this information in order to see how people perceive the company's

[1] Netcraft. Web Server Survey. August 2011. Netcraft | Internet Research, Anti-Phishing and PCI Security Services. 7 August 2011 <http://news.netcraft.com/archives/category/web-server-survey/>.
[2] Keyword for the purpose of this paper should be understood as search query, which may contain one or more keywords.



Statistical Challenges in eCommerce Research 2012

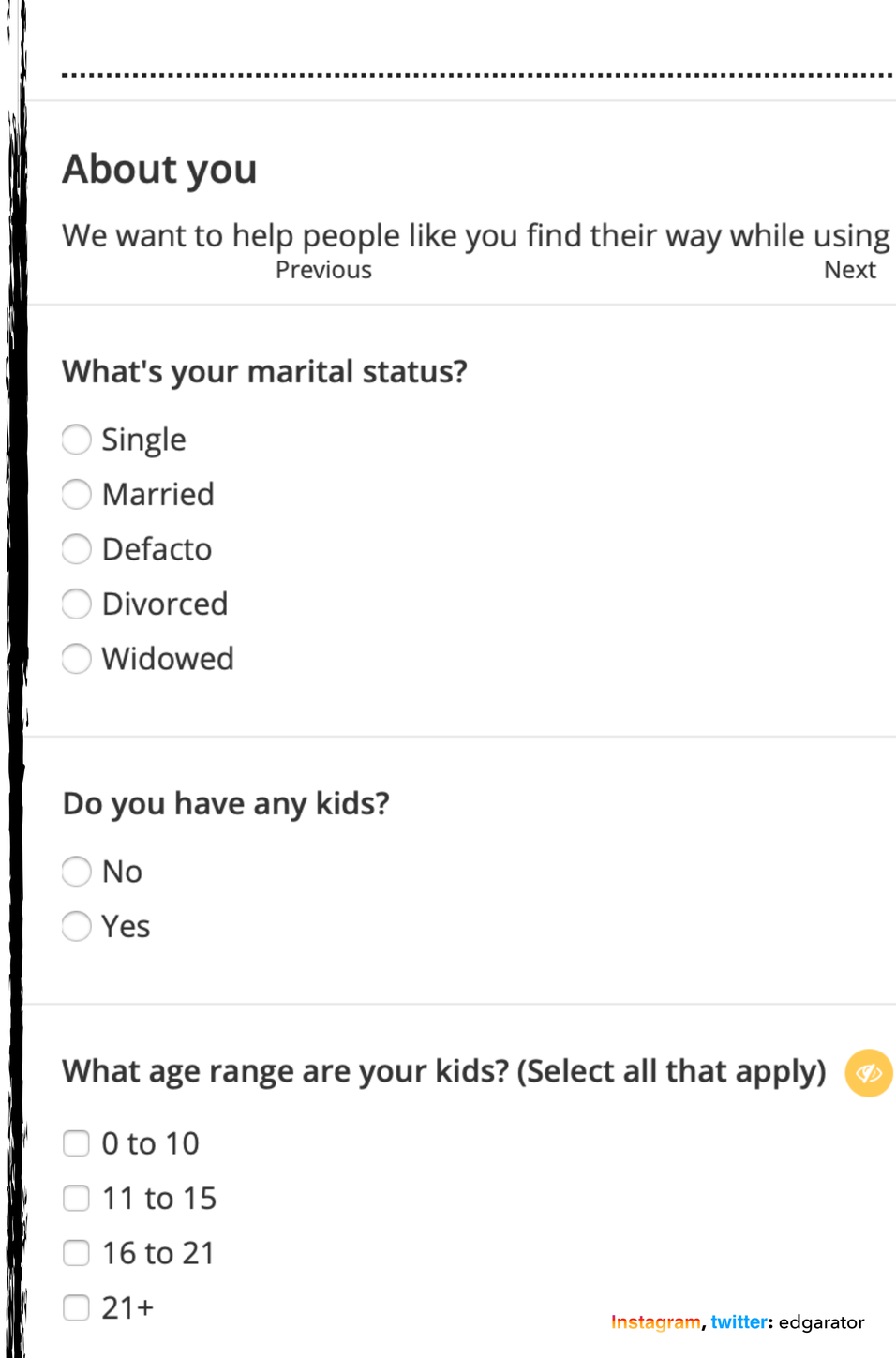
Anzaldua Moreno, E. and Esmailzadeh, R., 2012. Country Clustering Based on Search-Query Pattern Correlation. The Heinz Journal, 9(1), pp.5 - 11.

Tool: Gephi

If questions were **numeric** and **not categoric** we could do this easily!

Survey

- 20 Questions
- Closed survey
- Average of 5 options per question



Let’s quantify categorical answers

Quantifying Data

Binary answers

- Yes → 5
- No → 0

Quantifying Data

Mutually exclusive options:– i.e. Marital status, 4 options 🤷🏽

- Single → 0
- Married / De facto → 1.6667
- Divorced → 3.3333
- Widowed → 5

Quantifying Data

Mutually exclusive options:– i.e. Marital status, 4 options

- Single → 0y/o → 0
- Married / De facto → 32.3 → 1.9
- Divorced → 42 → 2.47
- Widowed → 85 → 5

Quantifying Data

Bracketed Data
(Multiple options allowed)
– i.e. Kids Age

- 0 to 10 → 1 if ☑ ; 0 if 🔲
- 11 to 15 → 1 if ☑ ; 0 if 🔲
- 16 to 21 → 1 if ☑ ; 0 if 🔲
- 21+ → 1 if ☑ ; 0 if 🔲

	0 to 10	11 to 15	16 to 21	21	Volume of Responses	Adjusted to 5
No Kids	0	0	0	0	15%	2.88
Kids own car	0	0	0	1	23%	4.42
Learning to drive	0	0	1	0	26%	5
Kids in College and Learning to Drive	0	0	1	1	5%	0.96
…	…					
Rabbits	1	1	1	1	0%	0

Quantifying Data

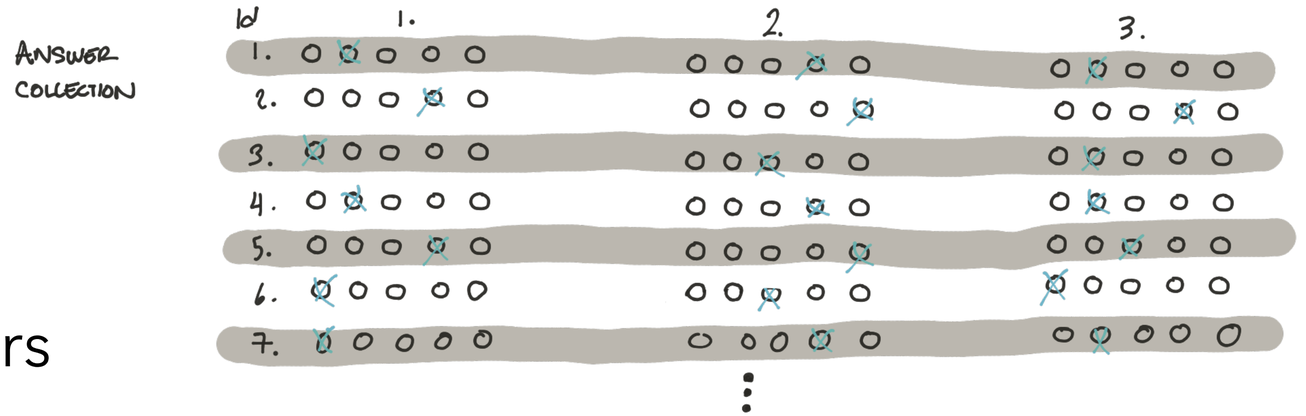
Mutually exclusive options:– i.e. Career stage, 7 options – Reframing in Years of Experience

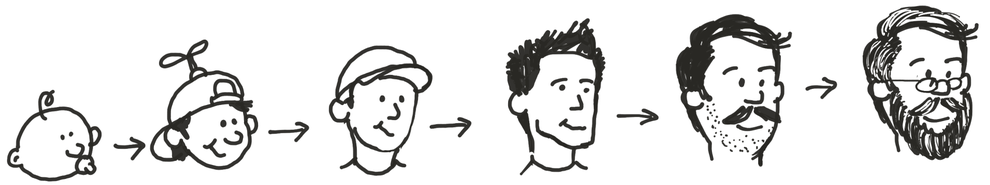
- Never worked → 0 → 0
- Junior → 3 → 0.5
- Mid-weight → 6 → 1
- Senior → 10 → 1.66
- Manager of others → 12 → 2
- Executive → 15 → 2.5
- Retired → 30 → 5

Quantifying Data

- The more we told the story with numbers the quality of the clusters improved.

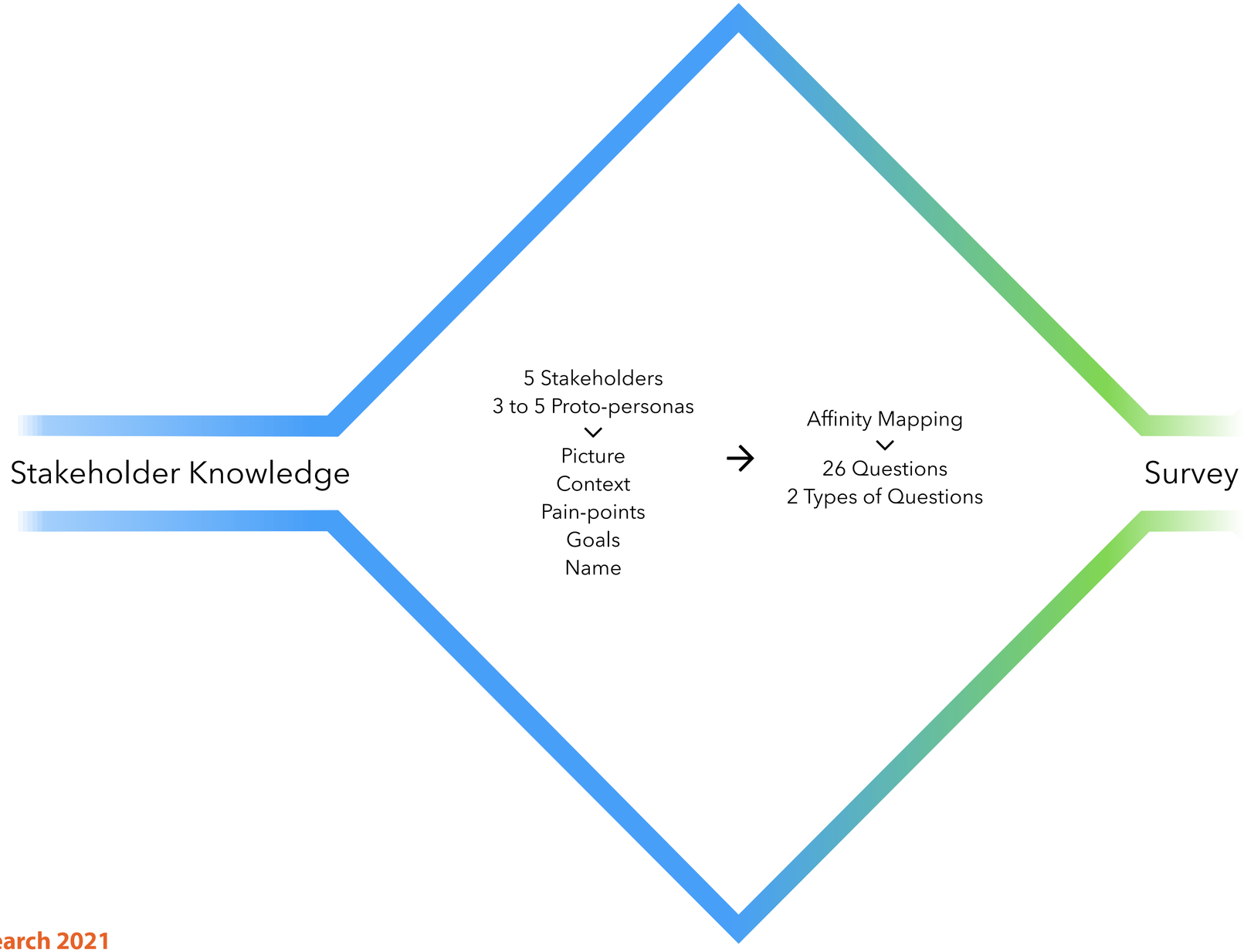
- Non-linear clusters were more useful





But how can I cluster the data?

Let's review the **cluster** process now!


Stakeholder Knowledge
5 Stakeholders
3 to 5 Proto-personas
Picture
Context
Pain-points
Goals
Name
Affinity Mapping
26 Questions
2 Types of Questions
Survey

rs
onas

Affinity Mapping

26 Questions
2 Types of Questions

Survey

Quantified
Responses

~800
responses

Python
Algorithm

Correlation Matrix
Community Detection

Response Clusters

Screeners

```
%matplotlib inline
import pandas as pd
import numpy as np
import networkx as nx
from datetime import datetime
from networkx.algorithms import community
from networkx.algorithms.community import greedy_modularity_communities

import matplotlib
import matplotlib.pyplot as plt
```

Algorithm A

Source	Target	Correlation
Response A	Response B	0.75
Response A	Response C	0.12
Response A	Response D	0.82
Response A	Response E	0.55
Response A	Response F	0.67
Response A	Response G	0.31
Response A	Response H	0.14
Response A	Response I	0.15
Response A	Response J	0.92

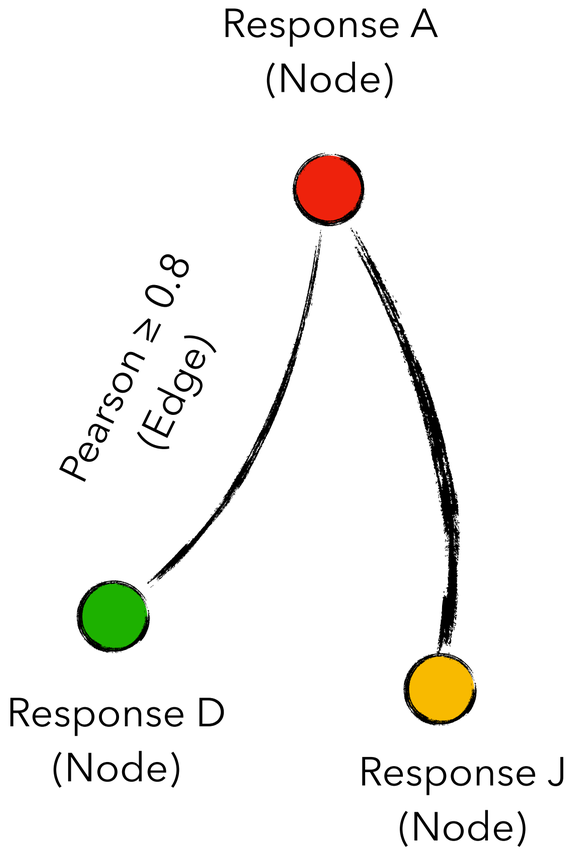
Arbitrary Correlation Value → 0.8

Algorithm A

Source	Target	Correlation
Response A	Response D	0.82
Response A	Response J	0.92
Response B	Response K	0.85
Response C	Response L	0.82
Response C	Response M	0.81
Response D	Response A	0.82
Response E	Response O	0.92
Response D	Response P	0.97
Response G	Response G	0.82
Response A	Response Q	0.93
Response R	Response I	0.89

Arbitrary Correlation Value → 0.8

Algorithm A



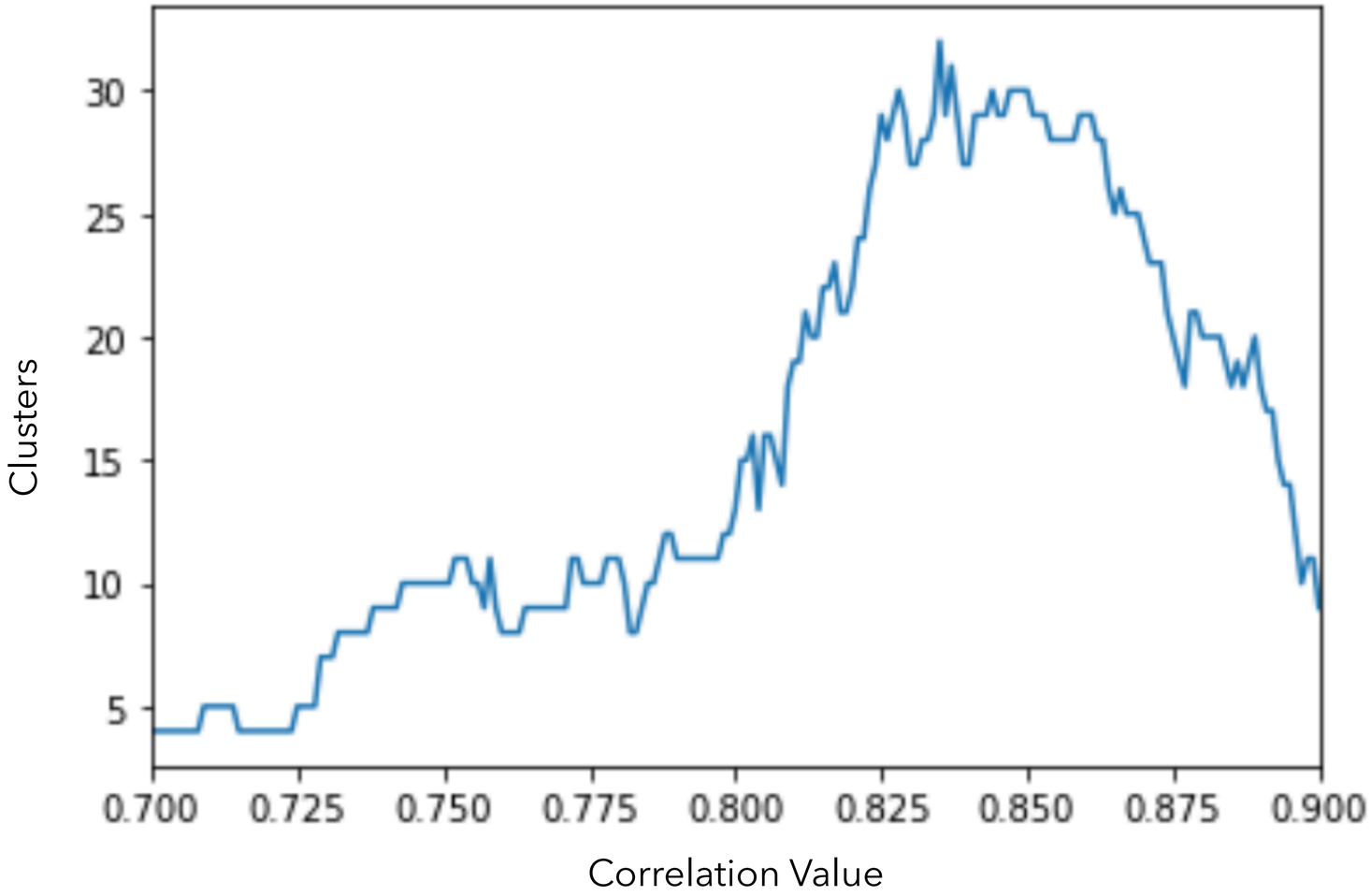
Source	Target	Correlation
Response A	Response D	0.82
Response A	Response J	0.92
Response B	Response K	0.85
Response C	Response L	0.82
Response C	Response M	0.81
Response D	Response A	0.82
Response E	Response O	0.92
Response D	Response P	0.97
Response G	Response G	0.82
Response A	Response Q	0.93
Response R	Response I	0.89

Arbitrary Correlation Value → 0.8

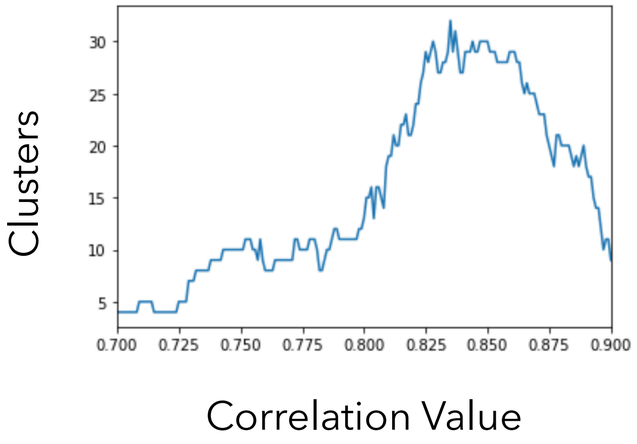
270 Nodes

~20 clusters

Algorithm B


Clusters
30
25
20
15
10
5
0.700
0.725
0.750
0.775
0.800
0.825
0.850
0.875
0.900
Correlation Value

Algorithm B

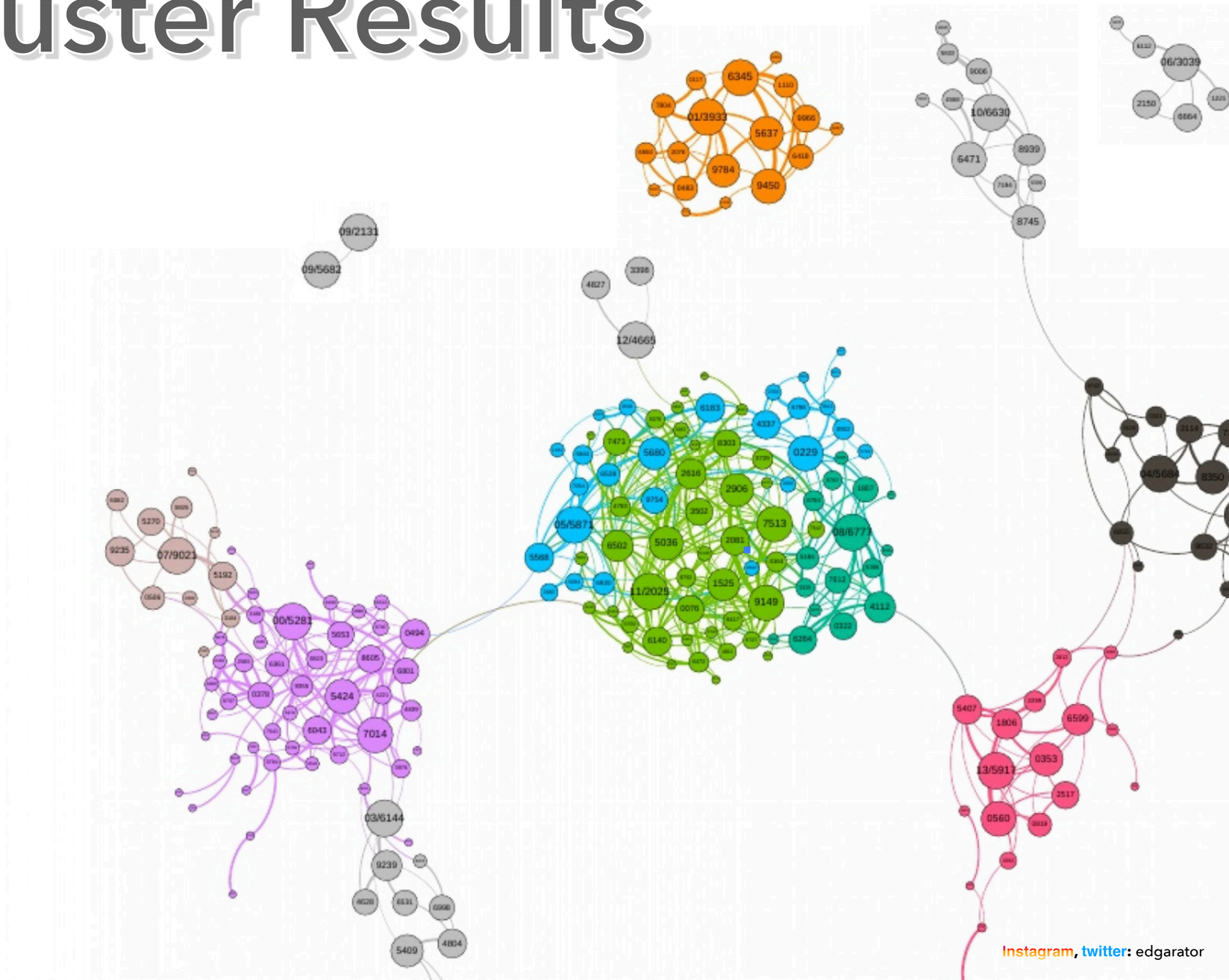


Correlation Value	Clusters	Responses
0.700	4	800
⋮		
0.780	15	390
0.785	**14**	**385**
0.790	16	384
⋮		
0.900	10	~100

Best Combination of

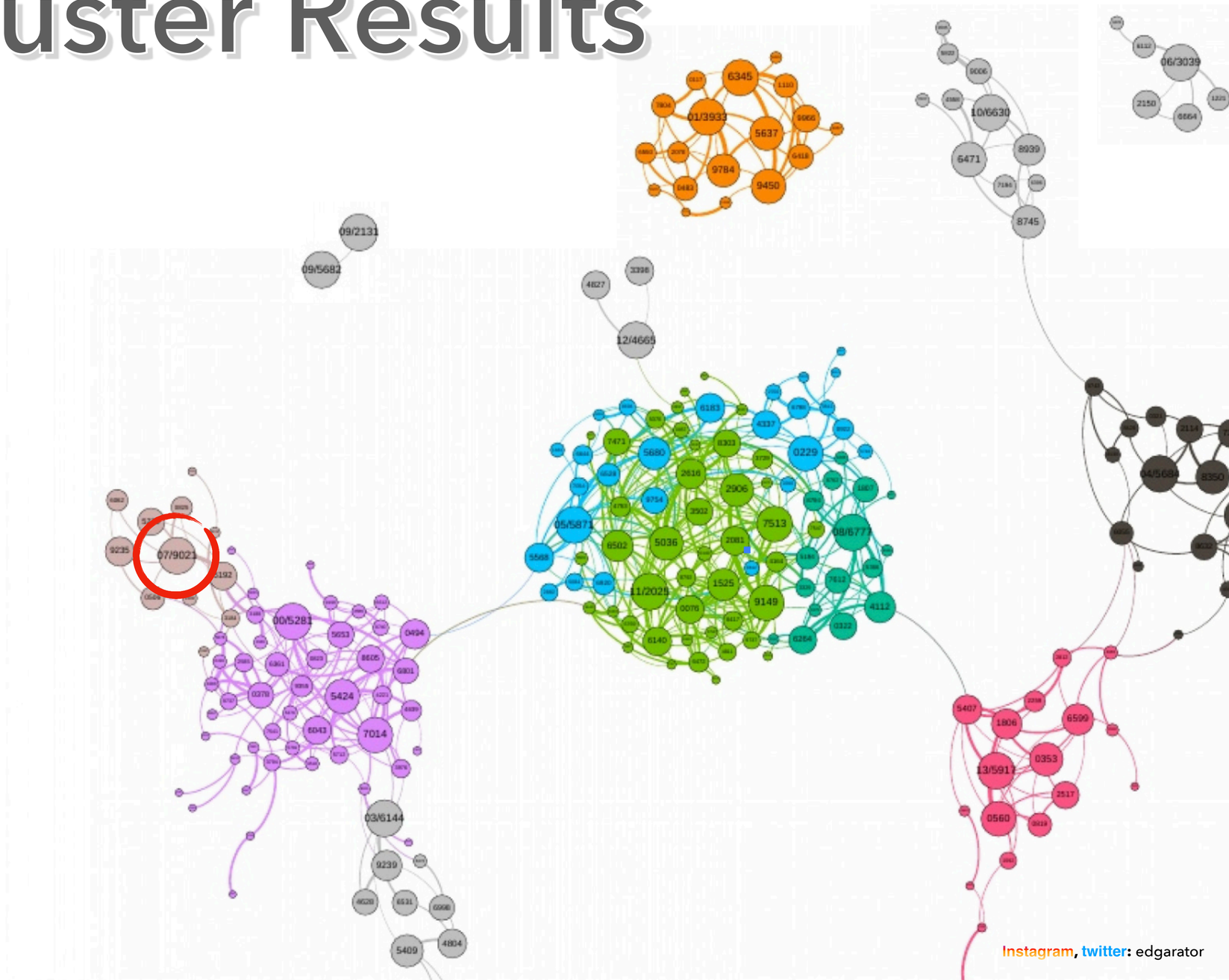
- Correlation (High)
- Least amount of clusters
- Enough responses included

Cluster Results



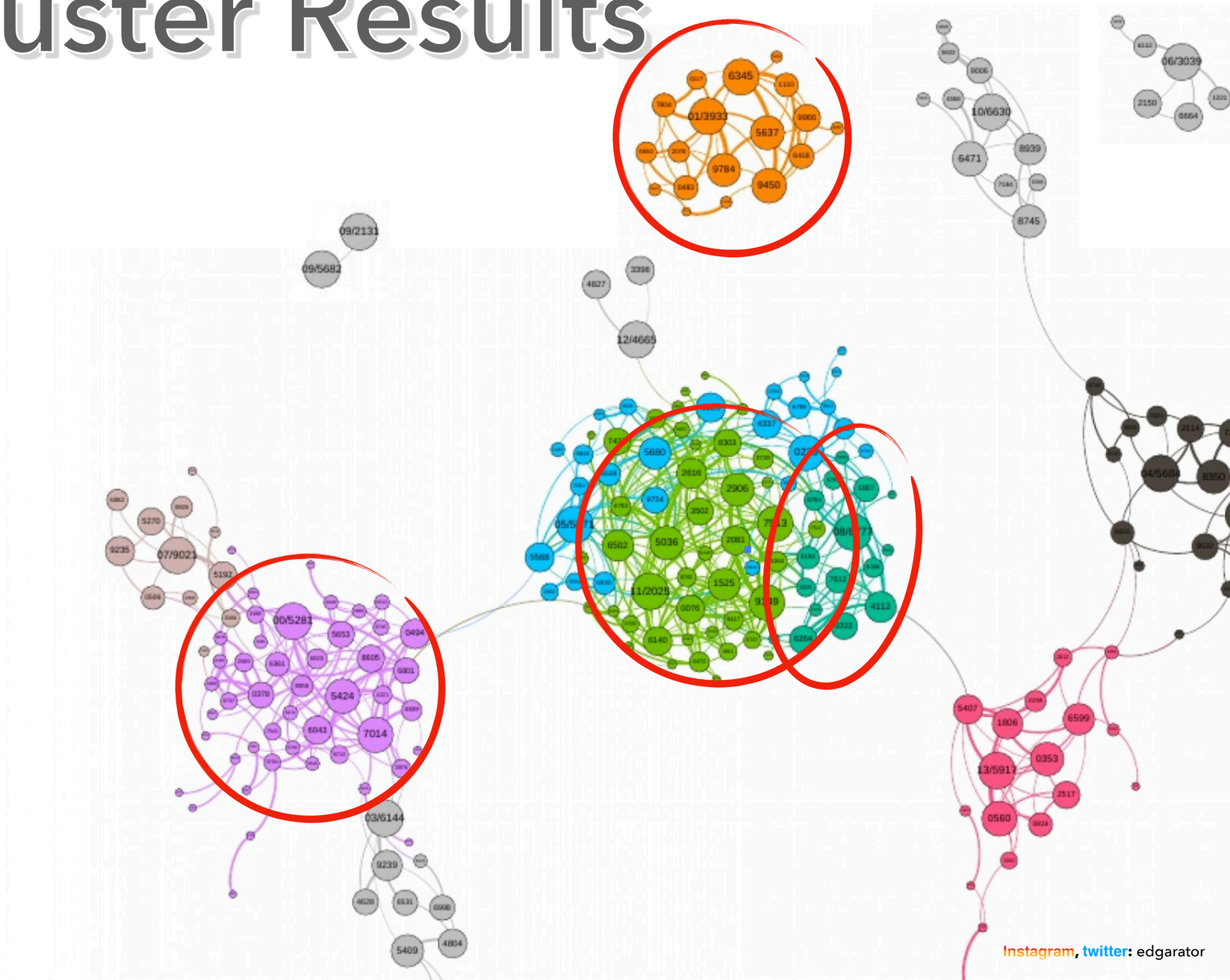
Cluster Results

- Survey response



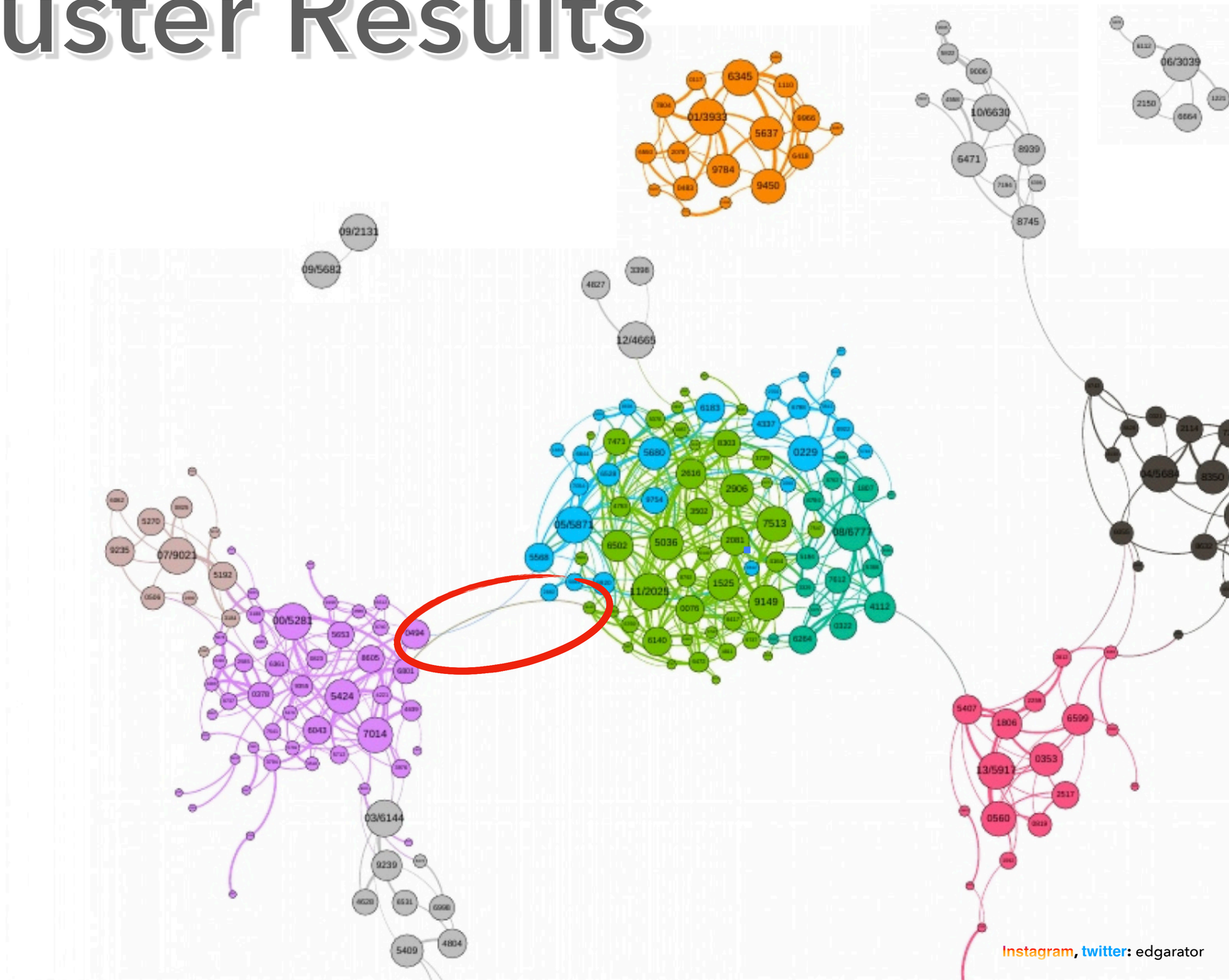
Cluster Results

- Survey response
- Clusters (by colour)



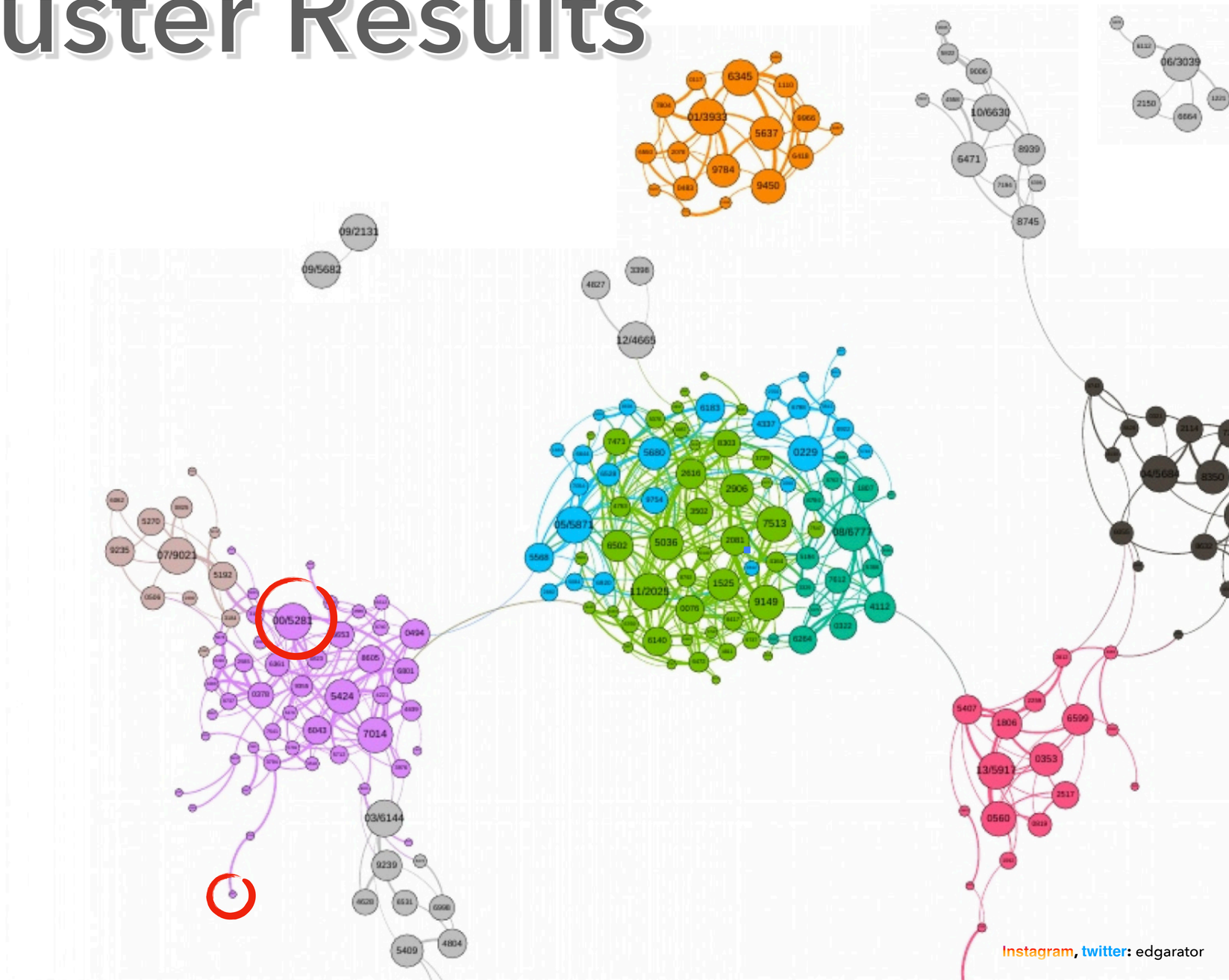
Cluster Results

- Survey response
- Clusters (by colour)
- Correlation degree



Cluster Results

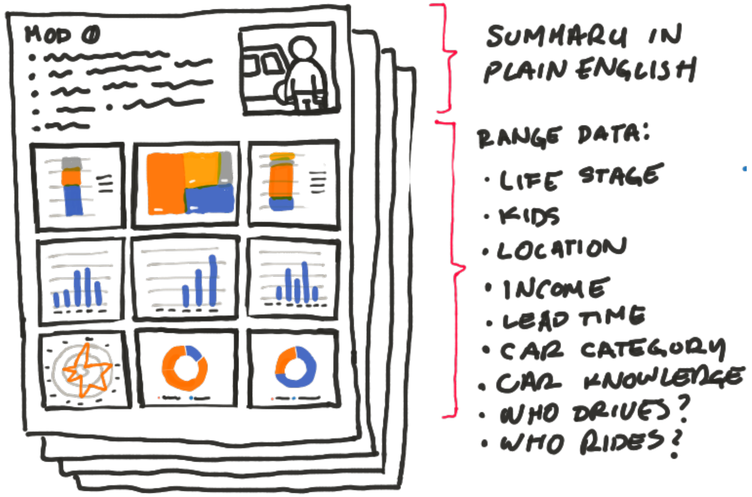
- Survey response
- Clusters (by colour)
- Correlation degree
- Response Centrality within Cluster



Cluster Results

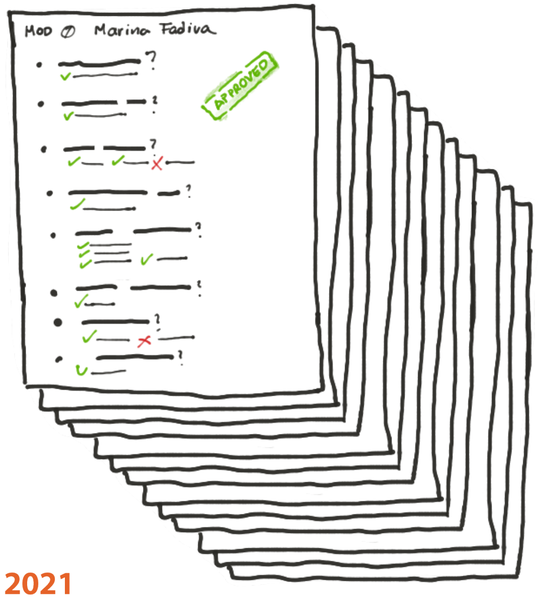
Each cluster could be read as

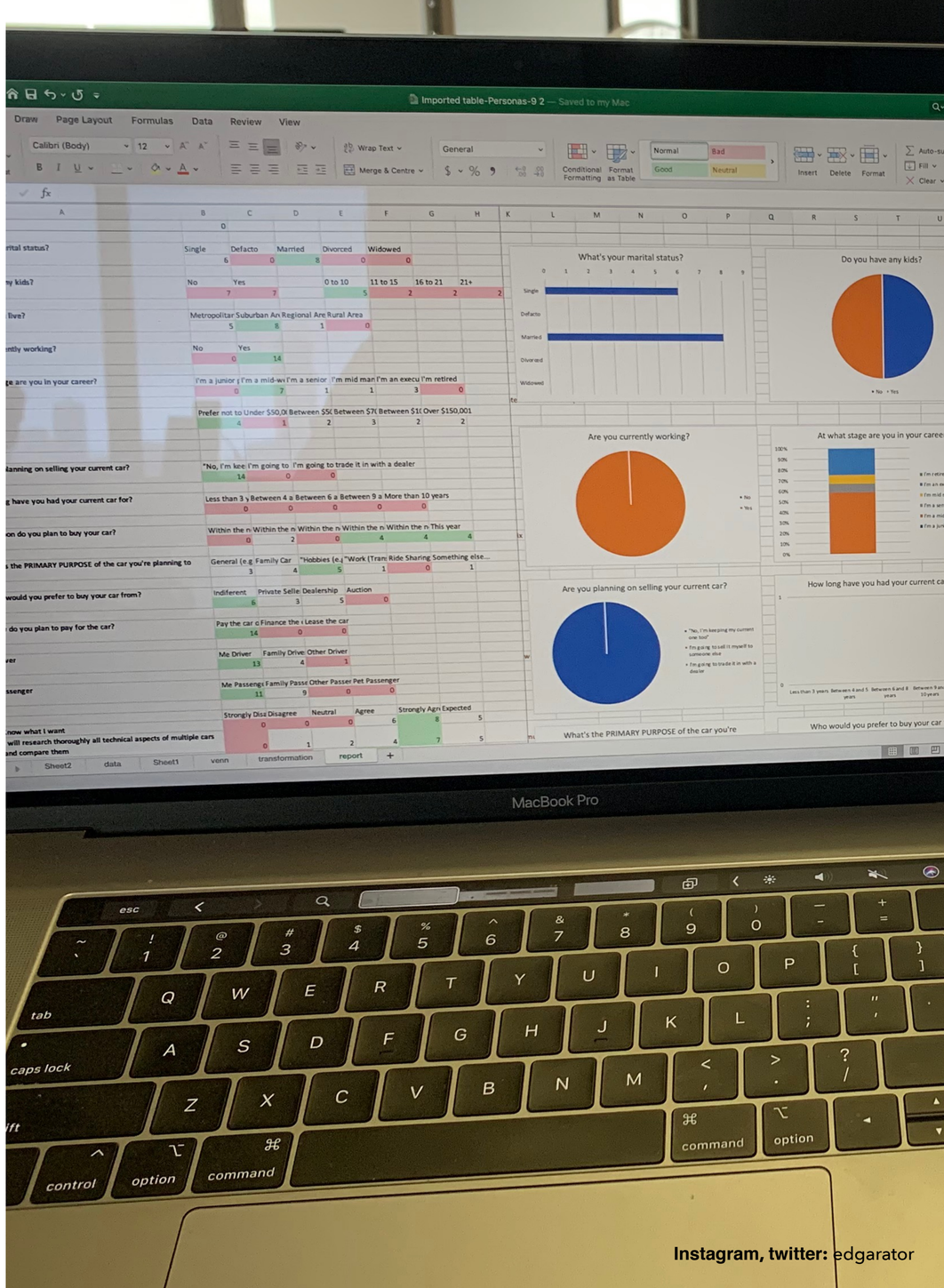
- <<marital status>> with <<No Kids/ Kids Age>> living in a <<Neighbourhood type>>.
- <<career stage>> (<<industry type>>) with a household income between <<household income range>>
- Etc…



Cluster Results

- ~~Personas~~
- ~~Segmentation~~
- 14 screeners for interviews

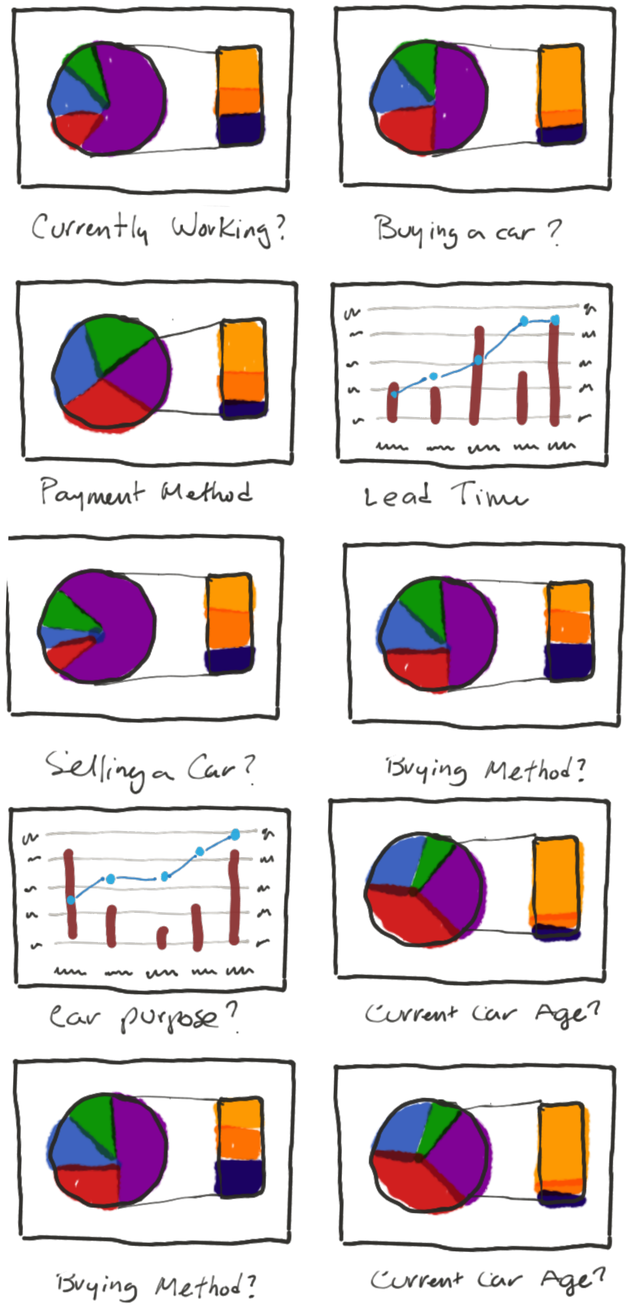




Instagram, twitter: edgarator

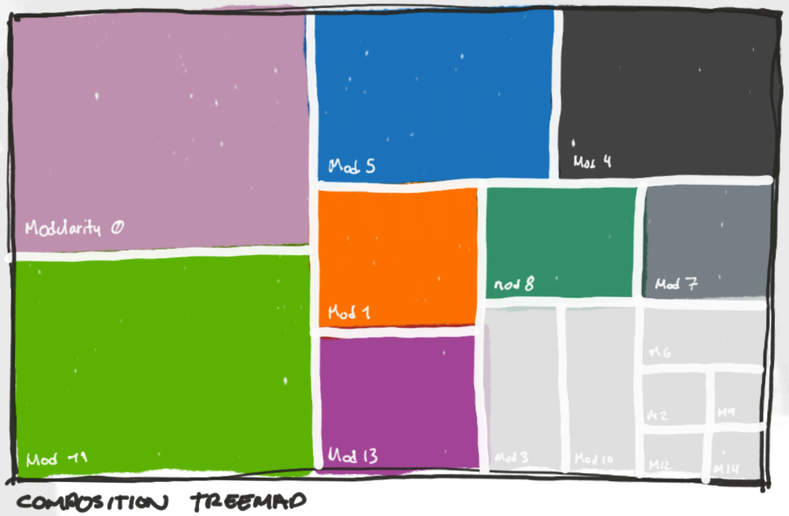
Bonus

- Statistically valid information about our user base
- Pack up and send as early findings – gets people excited :)

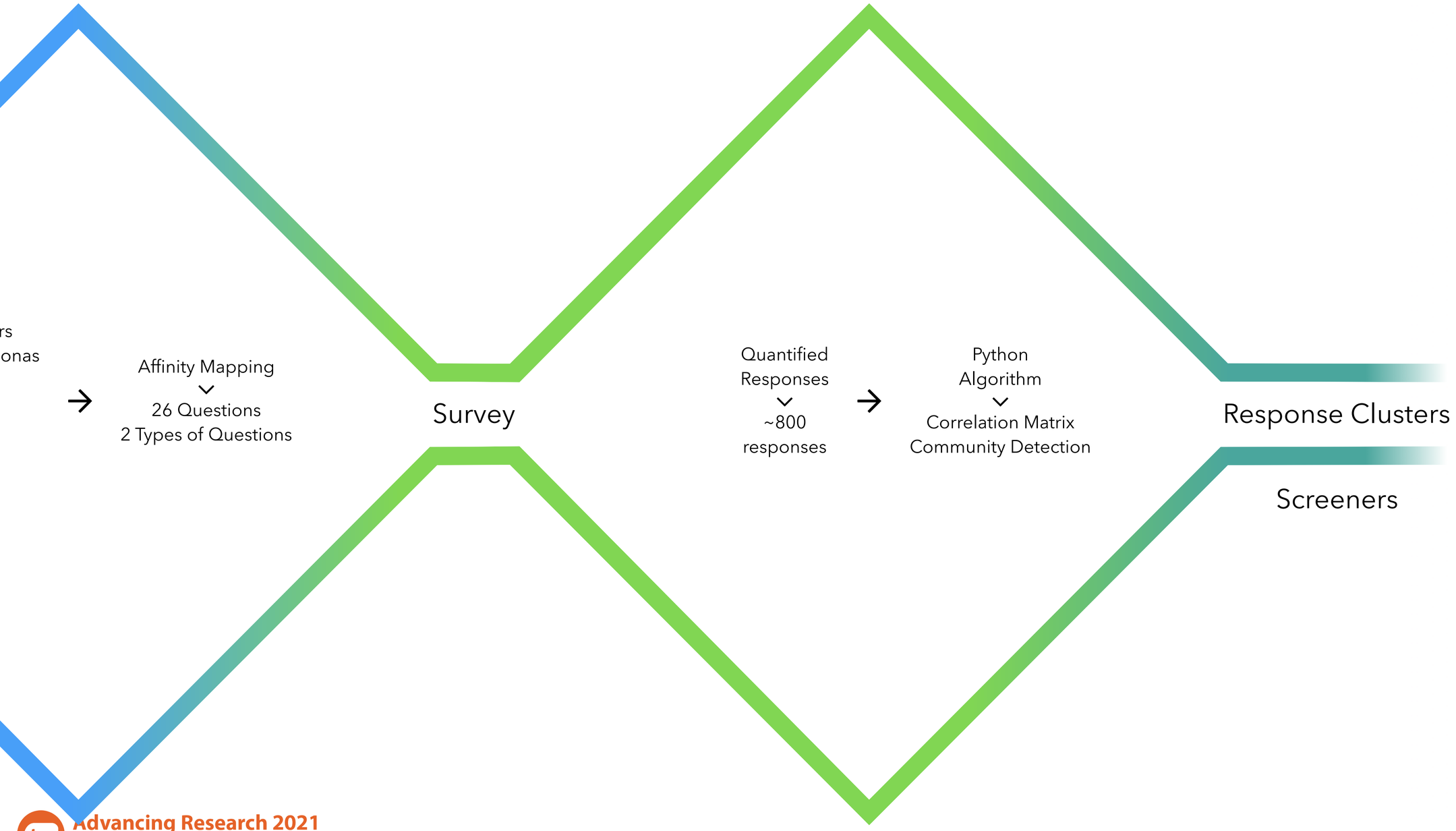


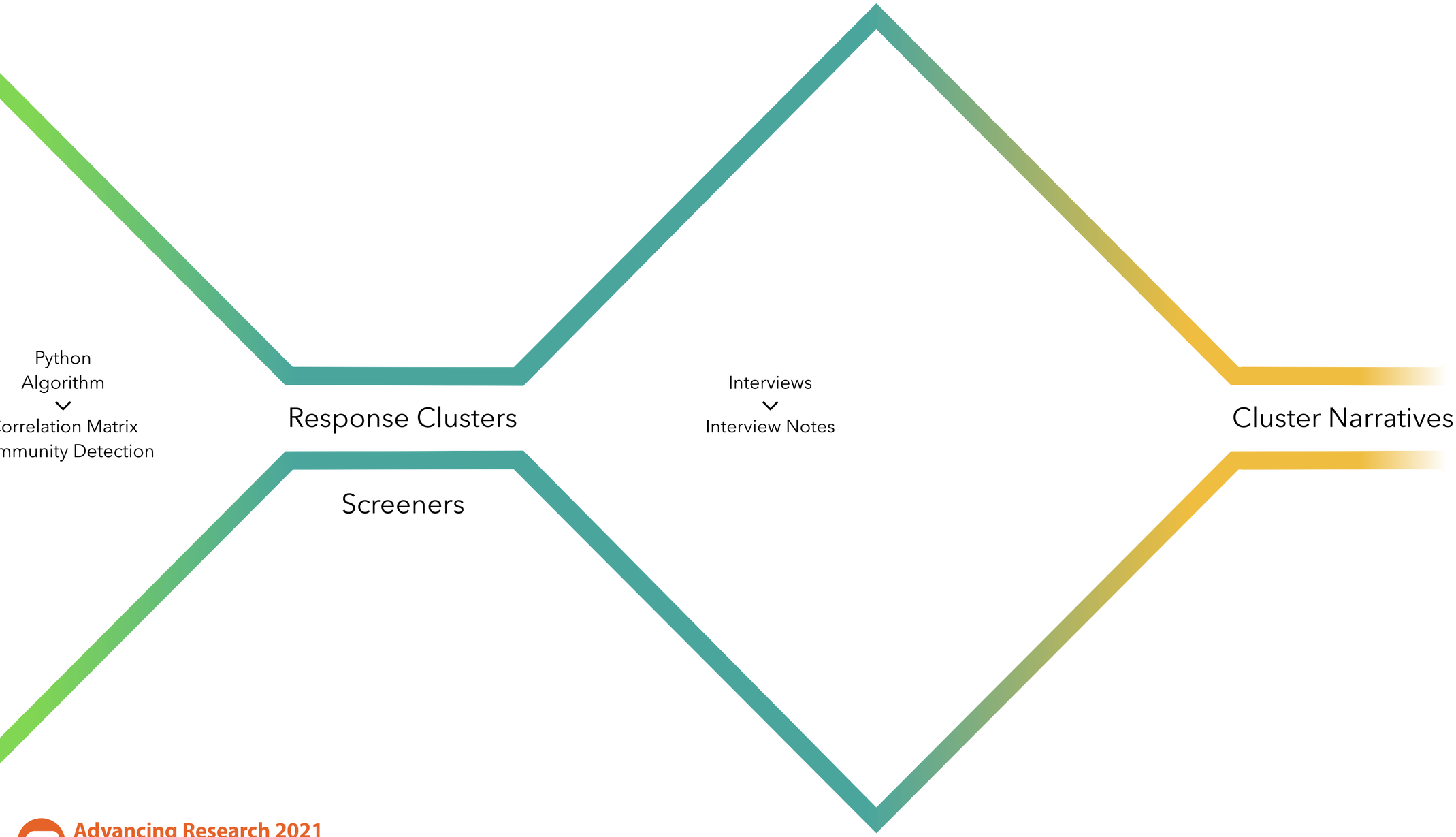
Interviews

RESPONSE NETWORK BY SIMILARITY



COMPOSITION TREEMAP


rs
onas
Affinity Mapping
26 Questions
2 Types of Questions
Survey
Quantified
Responses
~800
responses
Python
Algorithm
Correlation Matrix
Community Detection
Response Clusters
Screeners


Python
Algorithm
orrelation Matrix
mmunity Detection
Response Clusters
Screeners
Interviews
Interview Notes
Cluster Narratives

Advancing Research 2021
by Rosenfeld
From Research to UVP

Instagram, twitter: edgarator

Python
Algorithm
⌄
orrelation Matrix
mmunity Detection

Response Clusters

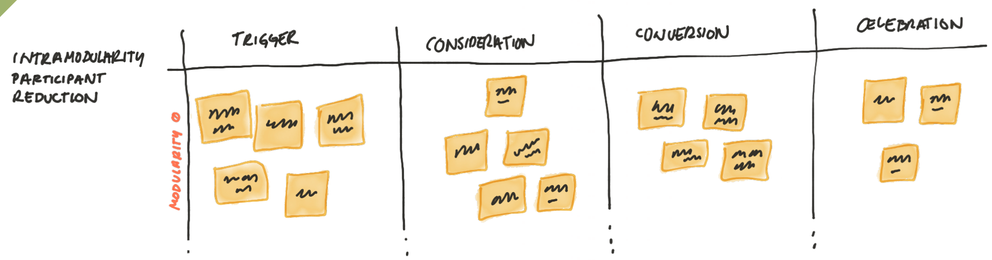
Screeners

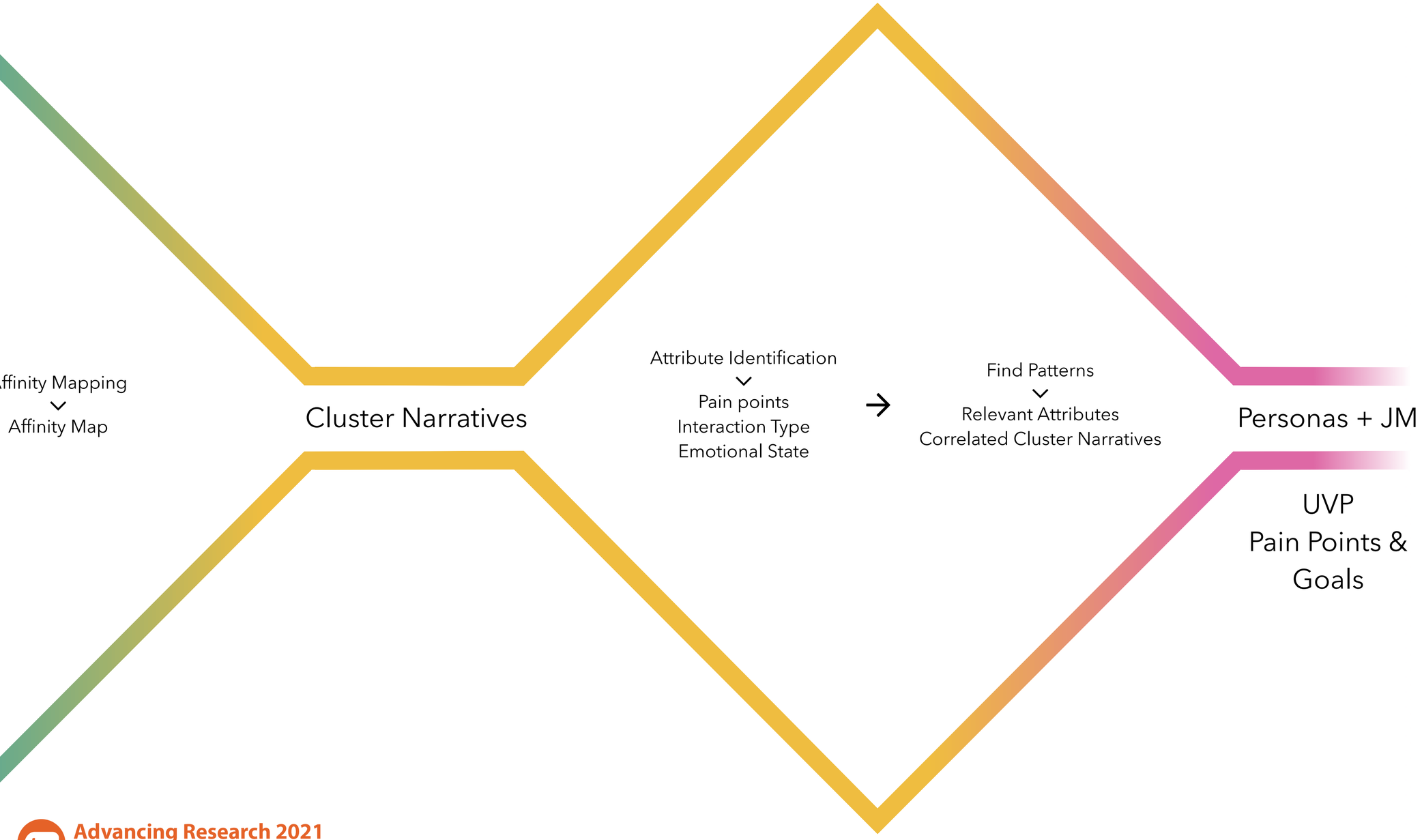
Interviews
⌄
Interview Notes

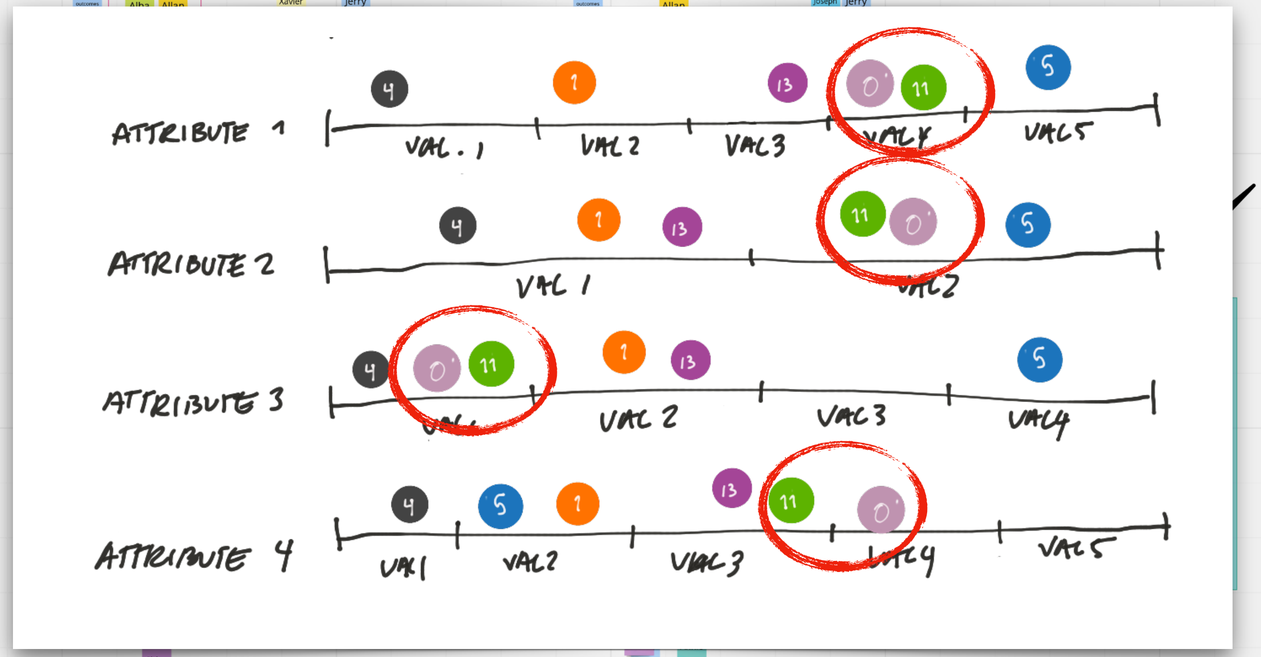
→

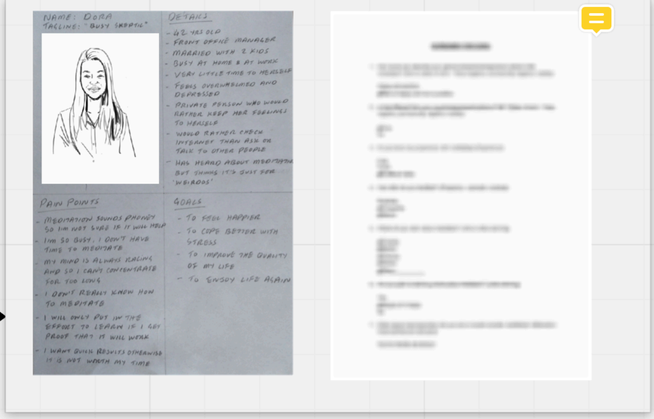
Affinity Mapping
⌄
Affinity Map

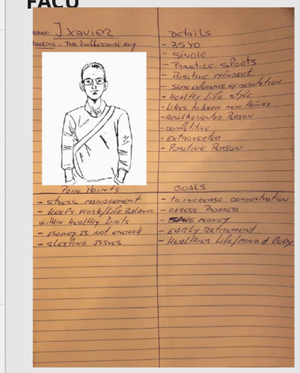
Cluster Narratives



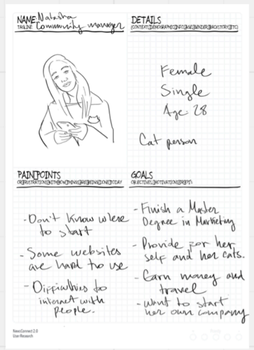

ffinity Mapping
Affinity Map
Cluster Narratives
Attribute Identification
Pain points
Interaction Type
Emotional State
Find Patterns
Relevant Attributes
Correlated Cluster Narratives
Personas + JM
UVP
Pain Points &
Goals







Paul



https://docs.google.com/document/d/1CAFc23QRKgJbUCR8X8BPxYHHN3v2lp_k5oLH2kzrG70/edit?usp=sharing

Advancing Research 2021
by Rosenfeld

Instagram, twitter: edgarator

ffinity Mapping

Affinity Map

Cluster Narratives

Attribute Identification

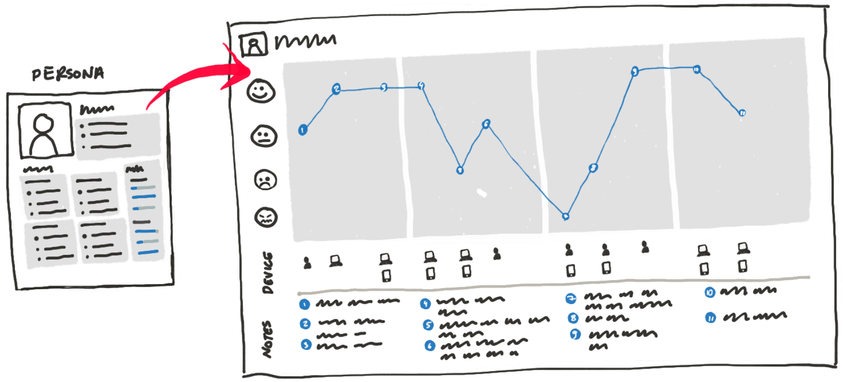
Pain points
Interaction Type
Emotional State

Find Patterns

Relevant Attributes
Correlated Cluster Narratives

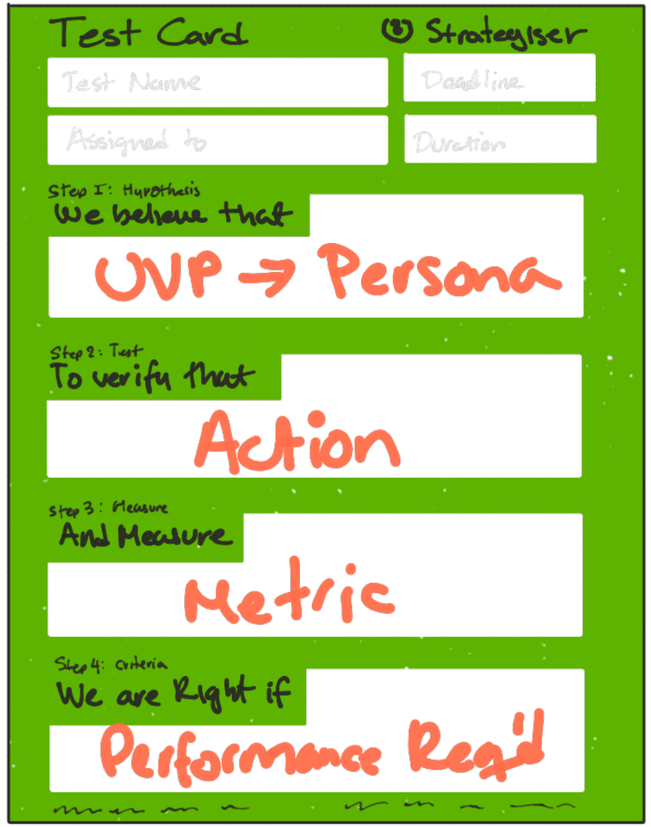
Personas + JM

UVP
Pain Points &
Goals



UVP

- Listing all pain-points and goals per persona
- Writing them as opportunity hypotheses
- Prioritising based on
  - Volume of lookalikes
  - Potential value to business
  - Ease of implementation



Source: https://www.strategyzer.com/hubfs/Assets%20-%20Downloads/the-test-card-1.pdf

ffinity Mapping

Affinity Map

Cluster Narratives

Attribute Identification

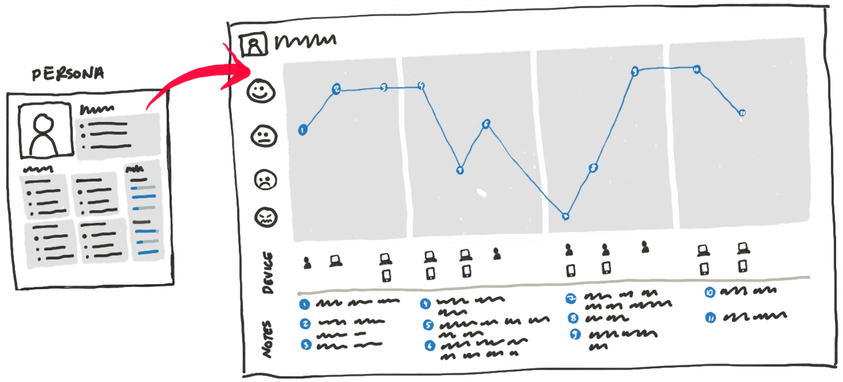
Pain points
Interaction Type
Emotional State

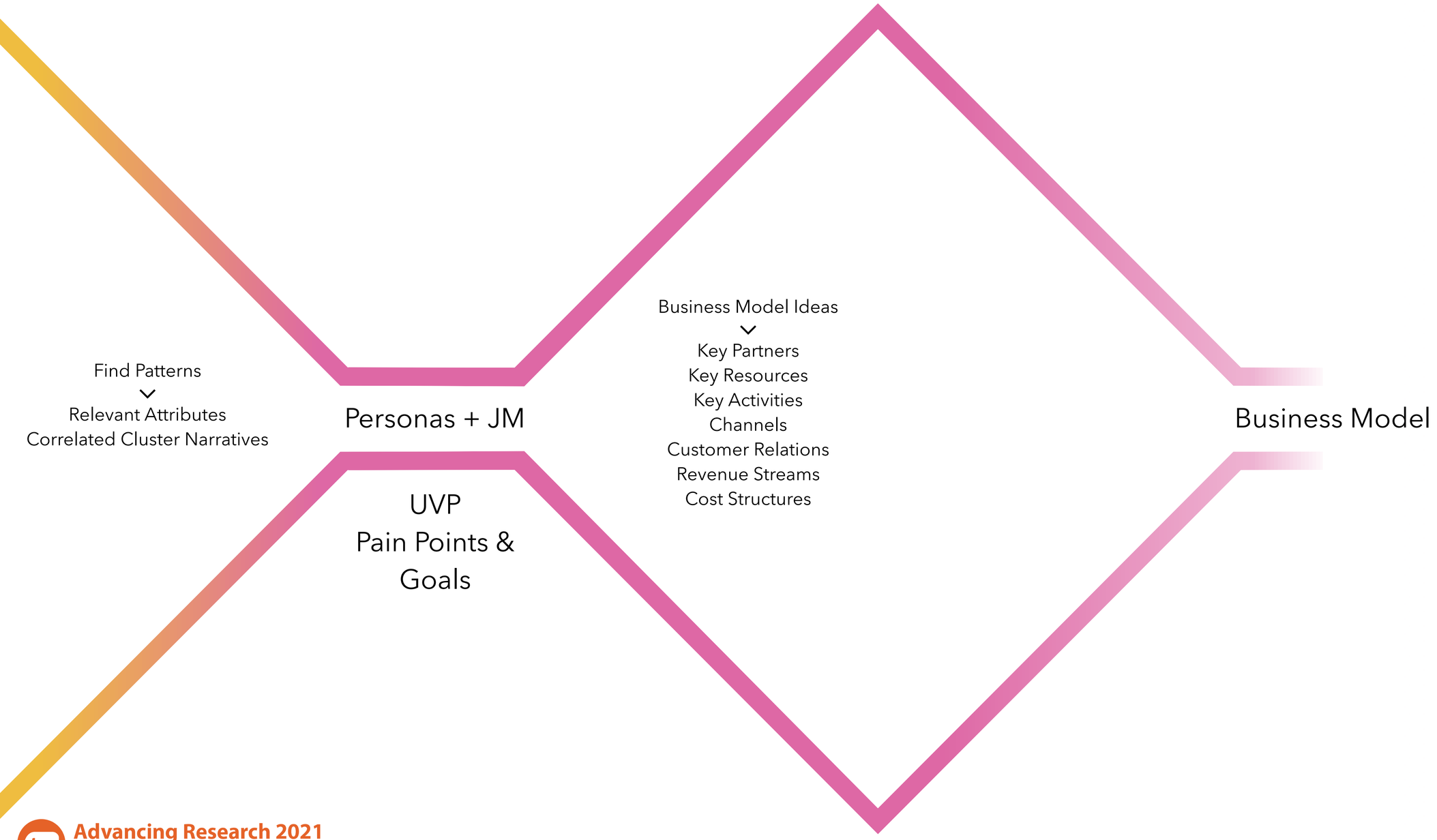
Find Patterns

Relevant Attributes
Correlated Cluster Narratives

Personas + JM

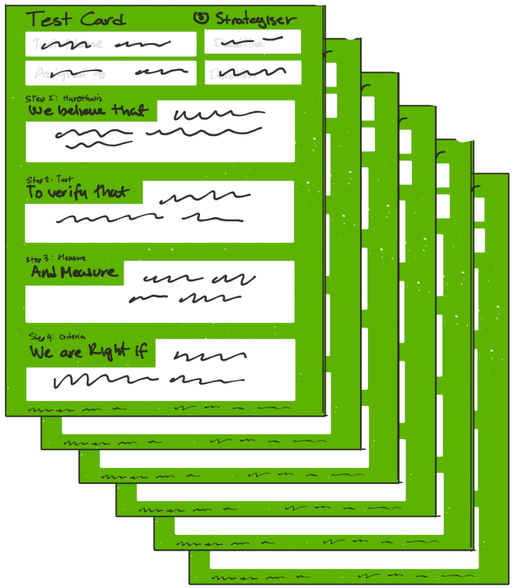
UVP
Pain Points &
Goals




Find Patterns
Relevant Attributes
Correlated Cluster Narratives
Personas + JM
UVP
Pain Points &
Goals
Business Model Ideas
Key Partners
Key Resources
Key Activities
Channels
Customer Relations
Revenue Streams
Cost Structures
Business Model

UVP

- Communicate as short stories
- Select what to focus on
- Figure out more than one way to deliver the UVP to the User Segment trying different revenue models and cost structures.



Find Patterns

Relevant Attributes
Correlated Cluster Narratives

Personas + JM

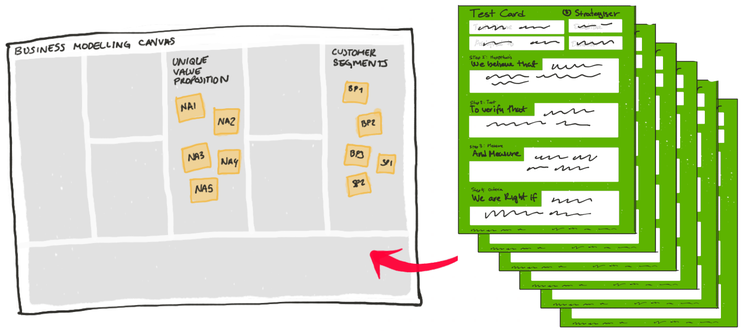
UVP
Pain Points &
Goals

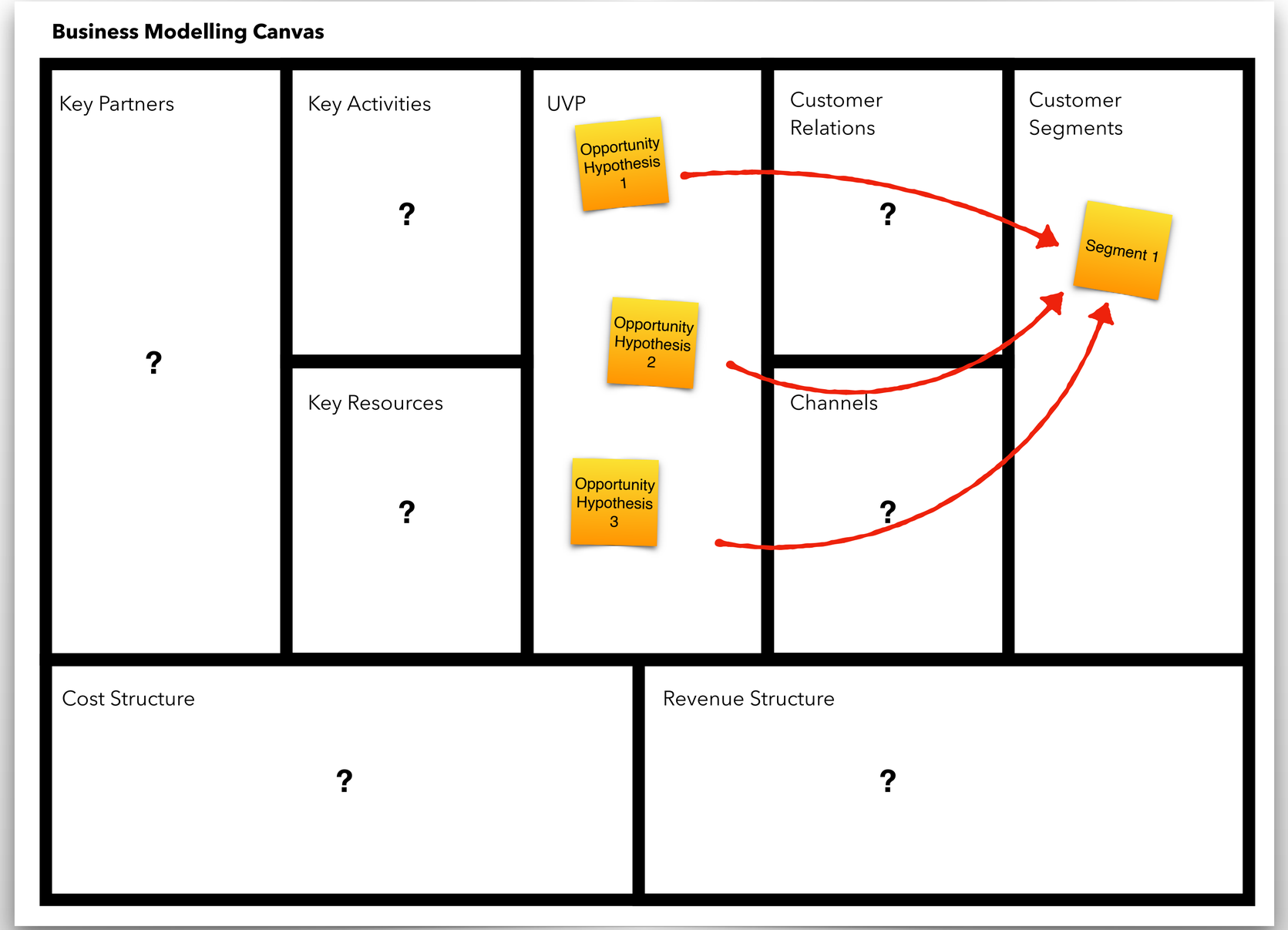
Business Model Ideas

Key Partners
Key Resources
Key Activities
Channels
Customer Relations
Revenue Streams
Cost Structures

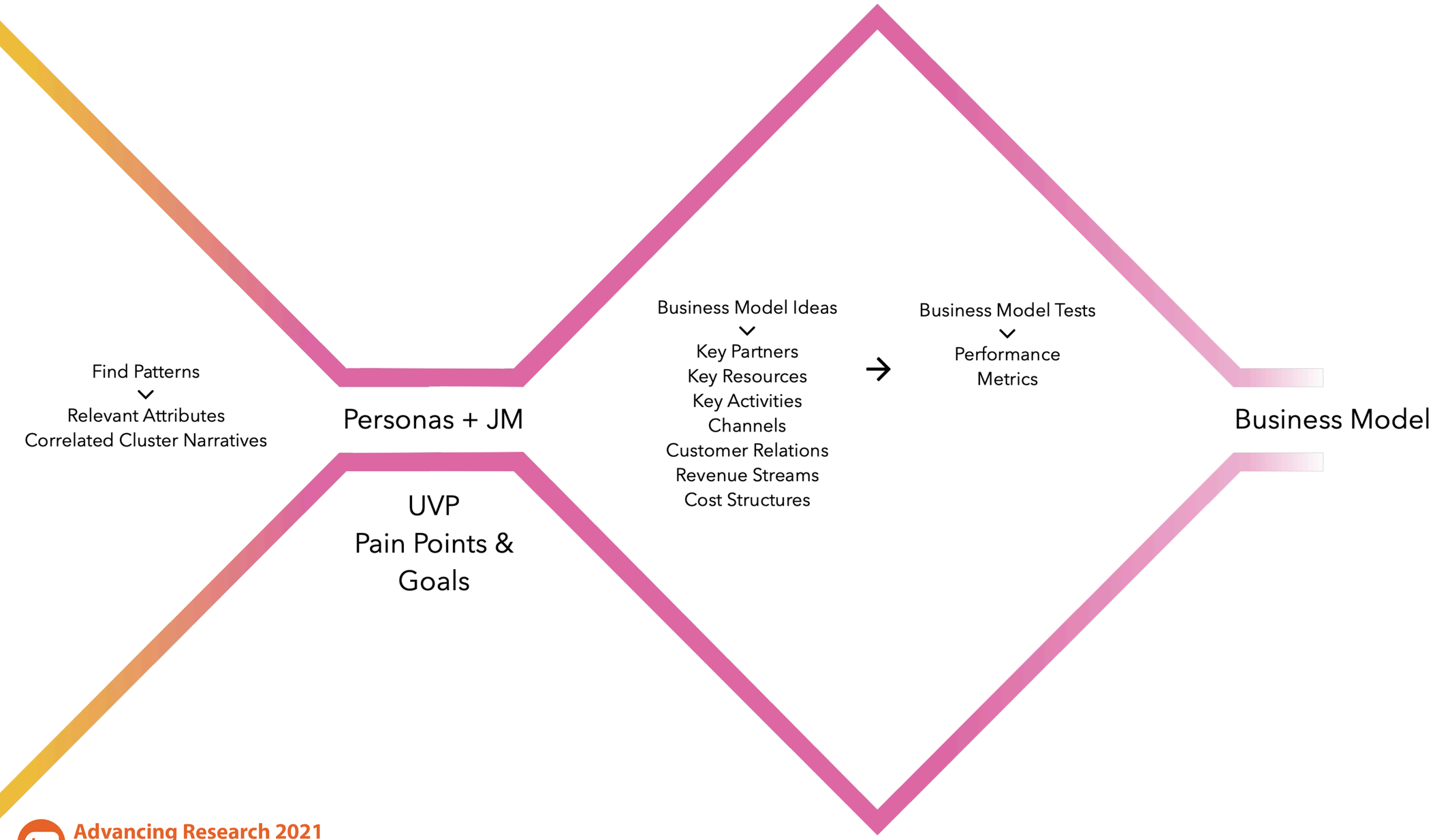
Business Model

NEEDS
VS.
CUSTOMER SEGMENTS
IN A BUSINESS
MODELLING CANVAS





Source: https://www.strategyzer.com/canvas/business-model-canvas


Find Patterns
Relevant Attributes
Correlated Cluster Narratives
Personas + JM
UVP
Pain Points &
Goals
Business Model Ideas
Key Partners
Key Resources
Key Activities
Channels
Customer Relations
Revenue Streams
Cost Structures
Business Model Tests
Performance
Metrics
Business Model

Summarising…

Quan Ask Frame Breakdown

Qual Interview Frame Breakdown

UVP List Frame Win

Wrap up

✓ Using design thinking to do research	→	**Design your as you design your products**
✓ Joining several research methods with a purpose in mind	→	**Don’t do research for research sake.**
✓ Focus on your users: stakeholders	→	**Tell them stories they understand, make them research-hungry**

Connect with me

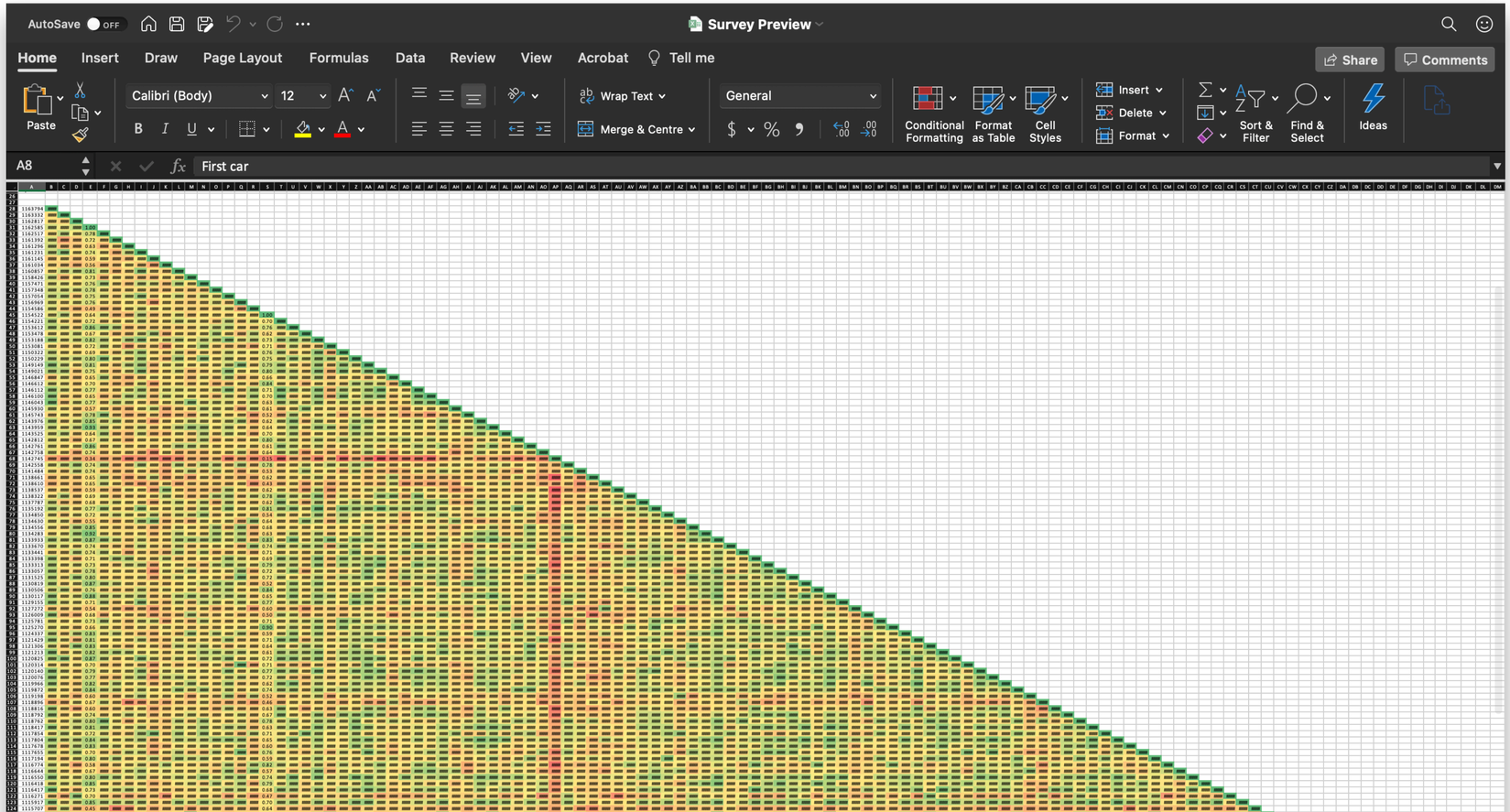
Linkedin http://www.linkedin.com/in/edgaranzaldua

http://edgar.design

Appendix

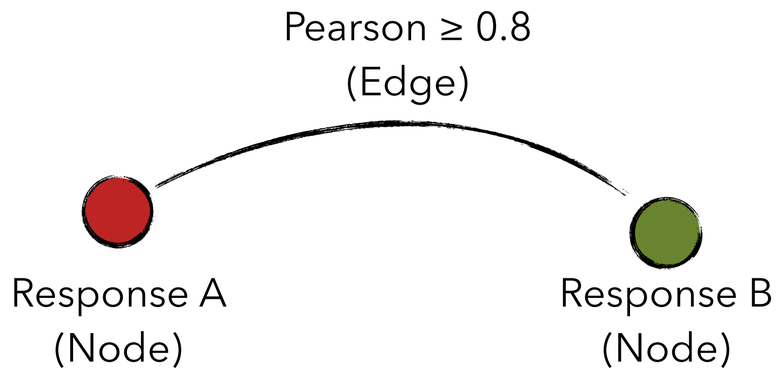
Algorithm A

- Compared each of 800 responses against all others → 640K Pearson Correlation calculations based on array of quantified answers



Algorithm A

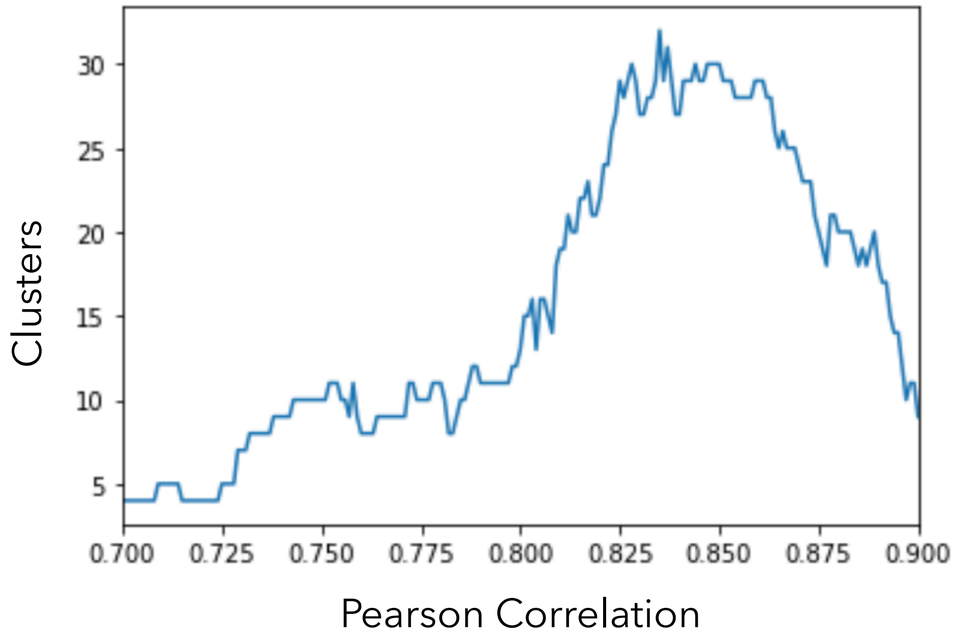
- Compared each of 800 responses against all others → 640K Pearson Correlation calculations based on array of quantified answers
- Picked an arbitrary correlation level → 0.8 *(perfect match 1)*
- Filtered out all response comparisons below 0.8
- Plotted network with:
  - Survey Responses as Nodes
  - Correlation Value as Edge



- First plot resulted in only 270 responses and around 20 clusters

Algorithm B

- Compared each of ~800 responses against all and filtered out anything under 0.7
  - Only three clusters
  - Almost all nodes included
- Increased the correlation threshold by 0.005 and repeated until reaching 0.9
- Correlation level 0.785 had:
  - The highest correlation value, with
  - the least of clusters present (14)
  - and 385 responses.



First...

- Have you ever seen **personas**...
  ...**done** properly, and/or...
  ...**used** properly

- Does your **research**
  ...always end **stored** in a file folder somewhere...
  ...and **seldom** get **used**?

- Is your company **confused** about...
  ...**what is it doing** and...
  ...**who is it doing it for**?

- You're a **data-nerd**