13F Holdings: Investor Clusters & Polarized Topic Detection

Capstone Project

RUI BAI, XINYI LIU, YICHI LIU, YUCHEN PEI, YUJING SONG

Columbia Data Science Institute

Student Group:
- Rui Bai (rb3454)
- Xinyi Liu (xl2904)
- Yichi Liu (yl4327)
- Yuchen Pei (yp2533)
- Yujing Song (ys3251)

Instructor:
- Adam S. Kelleher

Vanguard: IMFS and FIG

Project Sponsors:
- Maria Colangelo
- Min Fang

Mentors:
- Shadi Fadaee
- Flora Huang
- Stephen Lawrence

Overview of the Presentation

- Project introduction
- Cluster investors and understand their characteristics
- Identify cluster of investors that are similar/dissimilar to Vanguard
- Polarized topics detection methodology
- Polarized topics analysis
  - What type of investors drive the polarized topics
  - Which cluster drives the polarized topics
- Dashboard demo
- Q&A

Business Objective: Finding Insights from Holdings Data

Our investment desks want to use holdings data to understand how other market participants are approaching investments

- Understand how investors differ from Vanguard in their behaviors and holdings
- Identify emerging trends and deviations among investors
- Challenge: Holdings data is big and complicated

We can help analysts understand investor behaviors from 13F holdings data
- Understand how investors are clustered into groups, and what characteristics differentiate them
- Find investors that are similar to Vanguard and those that behave differently
- Identify instances of significant changes in holdings
- Represent insights with digestible visualizations

Project Motivation & Objectives

Motivation

- Understand institutional investors and their characteristics using 13F holdings data

- Identify characteristics of investors that differ from Vanguard

- Understand changes in investor behavior through time by looking at their portfolio holding changes

Objectives

- Understand institutional by creating a clustering model

- Understand how investors differ from Vanguard by analyzing the cluster that is dissimilar to Vanguard

- Detect significant polarized topics:
    - Identify the features of investors that are significant
    - Identify the cluster of investors that tends to drive the polarized topics

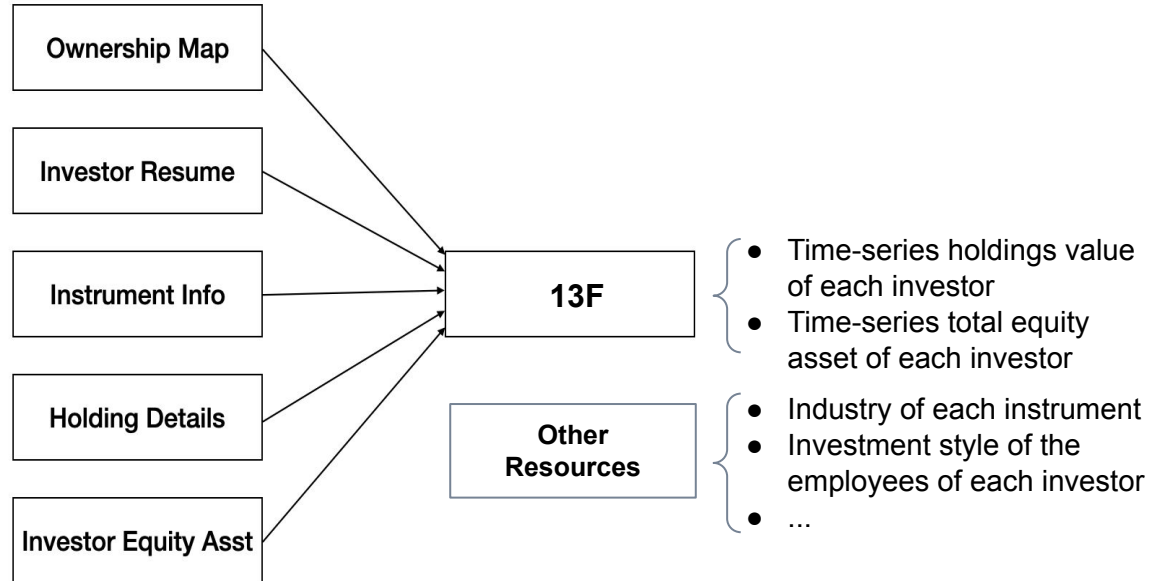# Data Overview – 13F Holdings (Source: Refinitiv Database)

13F Reports
- quarterly financial reports filed by institutional investment managers

Investors in our data
- 225 short-listed institutional investors

Data Scope
- Clustering: 2016Q2 - 2020Q2
- Polarized topics detection: 2010Q2 - 2020Q2

Ownership Map

Investor Resume

Instrument Info

Holding Details

Investor Equity Asst

13F

- Time-series holdings value of each investor
- Time-series total equity asset of each investor

Other Resources

- Industry of each instrument
- Investment style of the employees of each investor
- ...

Key Findings: Cluster investors into different groups to understand their characteristics

1 Cluster Investors that are similar/dissimilar to Vanguard and identify their characteristics

Cluster similar to Vanguard:

- Invest more in instruments with large & mega market capitalizations
- Invest more in the Software and IT service industry
- More total assets and more number of instruments
- Investment style has more focus on value

Cluster dissimilar to Vanguard:

- Invest more in instruments with small & middle market capitalizations
- Invest more in the Energy industry

- Less total assets and less number of instruments
- Investment style has more focus on emerging market

Key Findings: Understand differences in holding changes across different investors

2   Identify significant investment polarizations (polarized topics) at different time points

Polarized topics

- **Motivation**
  At a specific time, what are the differences in holdings changes for investors in a specific industry?

- **Definition**
  We define polarized topics as industries with large dispersion in their investors' position movements.

Time Point	Polarized Topics
2011 Q3	Utilities
2014 Q4	Energy, Real Estate
2015 Q3	Industrials, Technology, Utilities
2016 Q1	Utilities
2018 Q4	Financials
2020 Q1	Energy, Basic Materials, Consumer Cyclicals, Utilities, Real Estate
2020 Q2	Financials

Key Findings: Understand differences in holding changes across different investors and identify factors that drive these differences
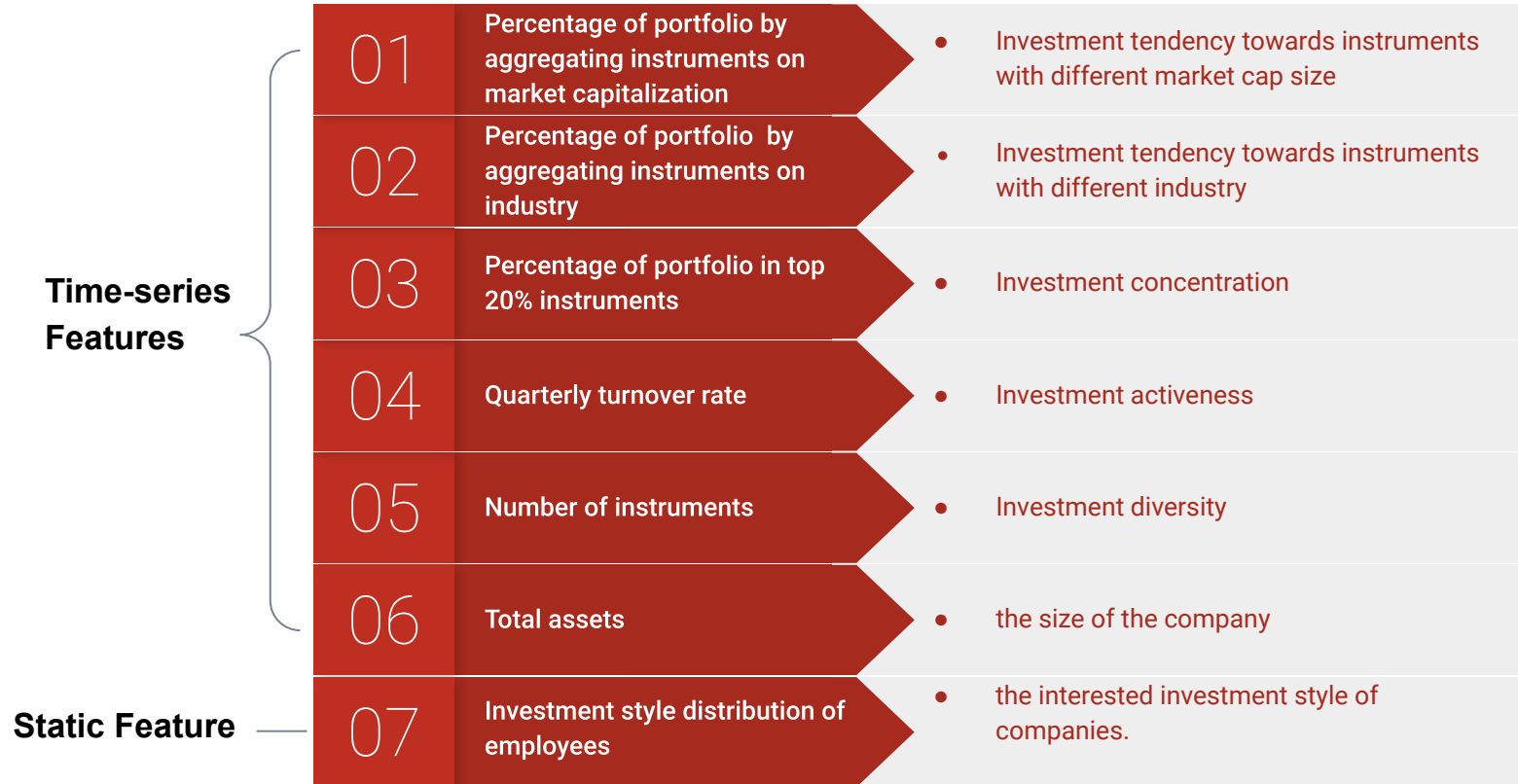
3 Identify characteristics that tend to drive the polarized topic

- From the perspective of investors cluster:
  - The investors that are dissimilar to Vanguard tend to drive the polarized topic

- From the perspective of investors features, these characteristics are present in polarized topics
  - High percentage portfolio investment in mid-cap instruments
  - High concentration on top 20% of its instruments
  - Fewer employees with the investment style of Index
  - Holding fewer positions
  - More employees with the investment style of Emerging Market

01 | Cluster investors and understand their characteristics

1 Seven Features of Investors: These features define clusters of investors

Time-series Features

01	Percentage of portfolio by aggregating instruments on market capitalization	● Investment tendency towards instruments with different market cap size
02	Percentage of portfolio by aggregating instruments on industry	● Investment tendency towards instruments with different industry
03	Percentage of portfolio in top 20% instruments	● Investment concentration
04	Quarterly turnover rate	● Investment activeness
05	Number of instruments	● Investment diversity
06	Total assets	● the size of the company

Static Feature

07	Investment style distribution of employees	● the interested investment style of companies.

1 Seven Features of Investors

The percentage of portfolio time series by aggregating instruments on market capitalization
- Market capitalization measures how much a company is worth

The percentage of portfolio time series by aggregating instruments on industry
- Classification of Instruments into 11 sectors and 30 categories

The percentage of portfolio in top 20% instruments time series
- This feature indicates how concentrated investors are in their top 20% instruments
  - Some Investors tend to concentrate on few instruments, others prefer distributed investments.
  - Higher Proportion of Top 20% Instruments -> "Concentrated Investment"

1 Seven Features of Investors

The quarterly turnover rate time series
- Turnover rate is the percentage of portfolios that have been replaced in a given period of time.
  - Higher turnover rate indicates more active investments
  - We take one quarter as our time slot

The number of instruments time series
- Describe whether the investor held many kinds of instruments.

The total assets time series
- Total assets indicate the size of the company.

Investment style distribution of employees (Static)
- The distribution of the employees that invest on different types of instruments could represent the investment style of the institutional investor.
- We take the proportion of employees with a specific investment style as one of the characteristic features.

02 | Identify cluster of investors that are similar/dissimilar to Vanguard

2.1 Investors Similar/Dissimilar to Vanguard

💡 By clustering with the 7 defined features, we identify a cluster with Vanguard and one without
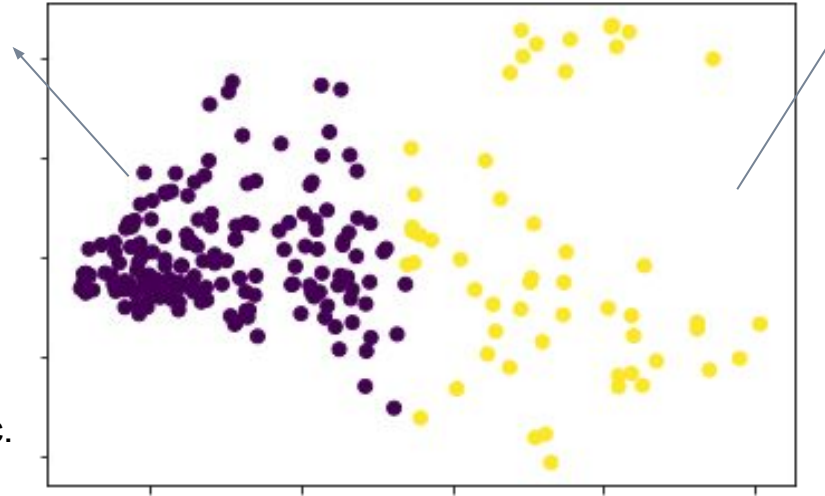
Cluster A:
171 investors
Similar to Vanguard

Examples:

- JP Morgan Asset Management
- UBS Financial Services
- Goldman Sachs
- The Vanguard Group, Inc.
- BlackRock Institutional

With Vanguard   **Without Vanguard**

Cluster B:
54 investors
Dissimilar to Vanguard

Examples:

- Apollo Capital Management, L.P.
- Angelo, Gordon & Co.
- New York Life Investment Management, LLC
- King Street Capital Management, L.P.
- Brookfield Asset Management, Inc.

2.2 Comparison of the Two Clusters

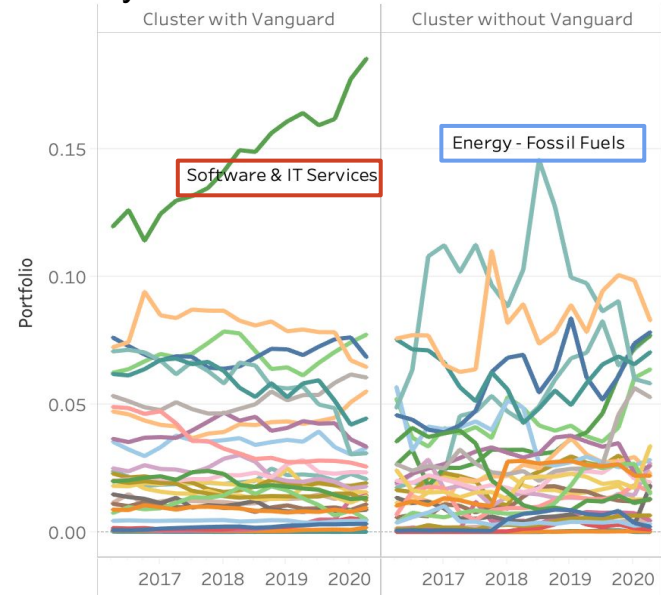**(1)** Tendency towards instruments with different market cap size
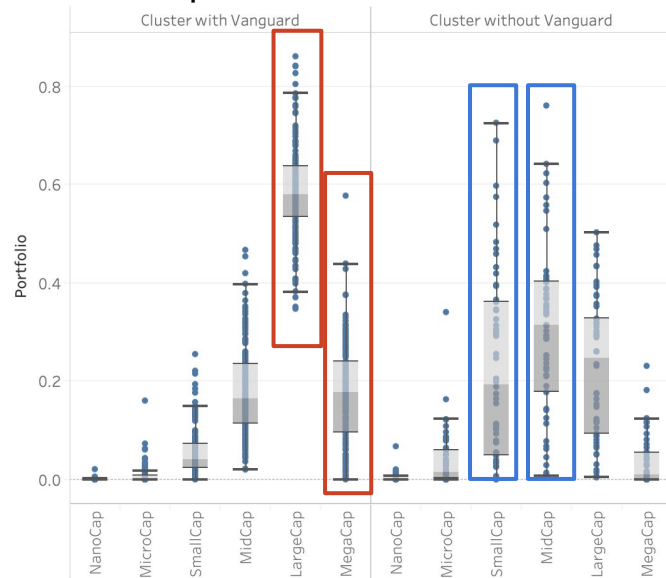
Cluster similar to Vanguard:
Invest more in instruments with large & mega market cap

Cluster dissimilar to Vanguard:
Invest more in instruments with small & middle market cap

(2) Tendency towards instruments with different industry

Cluster similar to Vanguard:
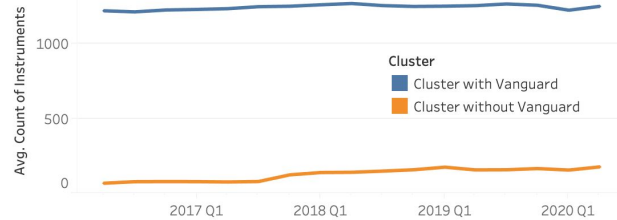Invest more in the Software and IT service industry

Cluster dissimilar to Vanguard:
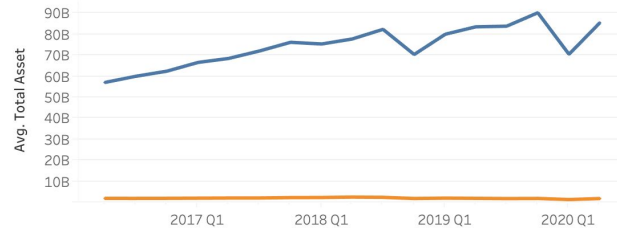Invest more in the Energy industry

2.2 Comparison of the Two Clusters

3 Size of the company

Average Count of Instruments Held for Each Cluster

Cluster
- Cluster with Vanguard
- Cluster without Vanguard

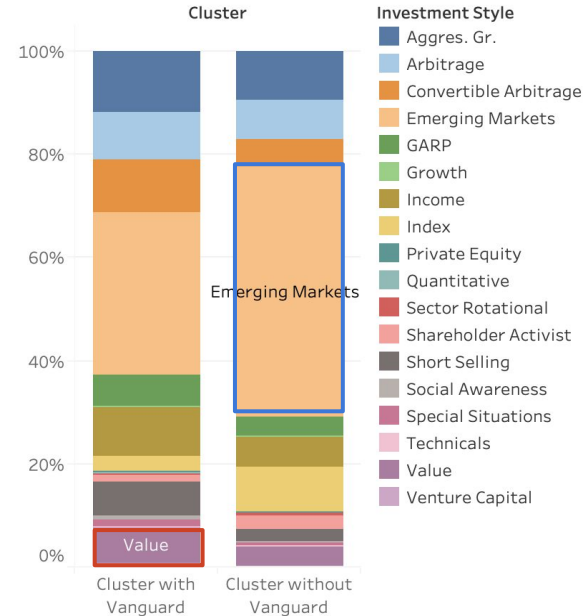### Average Total Asset Held for Each Cluster

Cluster similar to Vanguard:
More total assets and more number of instruments

Cluster dissimilar to Vanguard:
Less total assets and less number of instruments

4 Employees' investment style

Cluster

Investment Style
- Aggres. Gr.
- Arbitrage
- Convertible Arbitrage
- Emerging Markets
- GARP
- Growth
- Income
- Index
- Private Equity
- Quantitative
- Sector Rotational
- Shareholder Activist
- Short Selling
- Social Awareness
- Special Situations
- Technicals
- Value
- Venture Capital

Cluster similar to Vanguard:
Investment style focuses on value

Cluster dissimilar to Vanguard:
Investment style focuses on emerging market

03 | Polarized Topic Detection Methodology

Polarized topic: The industry in which investors have large dispersion of position movements

In an industry, use **delta** to represent the position movements, or more specifically, the change of percentage in an investor's portfolio holdings
- ➕ delta is positive
- ➖ delta is negative
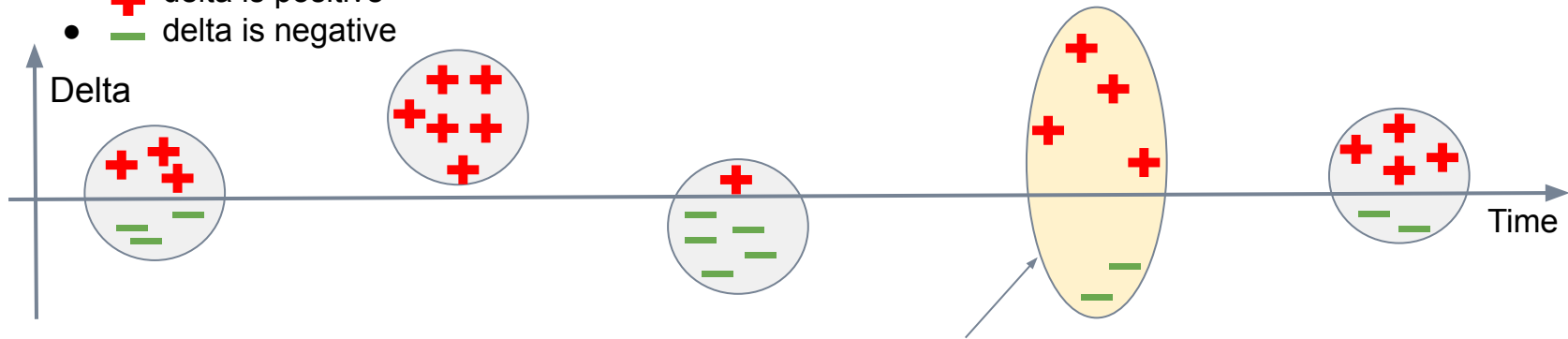
Delta

Time

*Each circle represents the investors' delta distribution at a certain time point

Large Dispersion (what we want to detect)

Large dispersion of deltas → Disagreement among the investors → **Polarized topic**

3.2 At each timepoint, identify the level of polarization in each industry
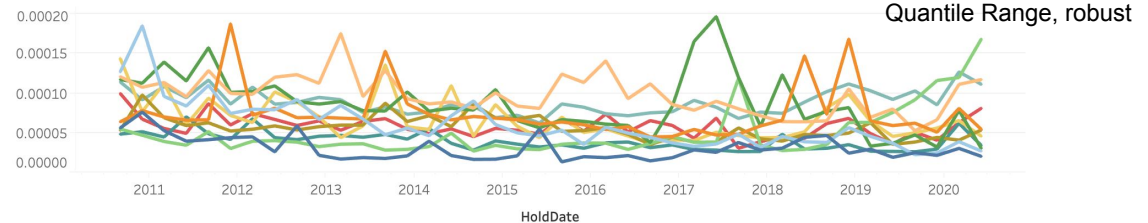
Spikes in plots show polarization

- We aggregate Instruments into 11 industries

- To find instances of significant polarizations, we look at the level of polarization among investors in each industry through time

- The spikes represent points in time where investors disagree with one another (i.e. polarized opinions) in a particular industry.
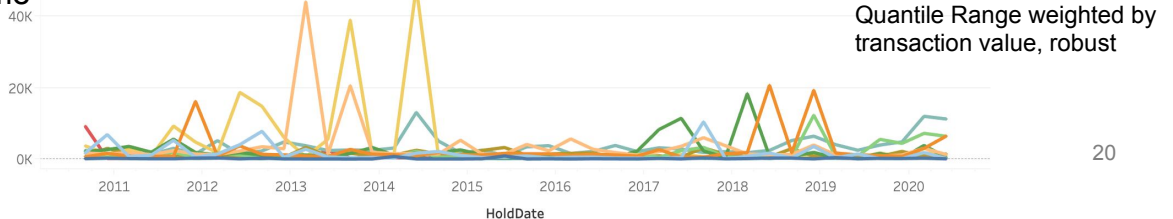
Polarization Measurement 1 (**Variance**) for Each Industry by Time

Variance, sensitive to extreme values
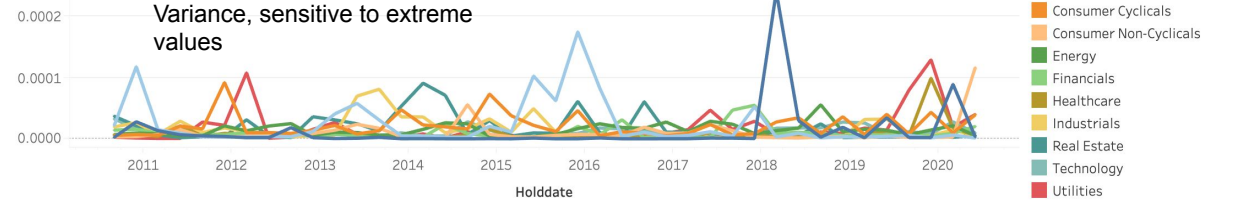
Polarization Measurement 2 (**IQR**) for Each Industry by Time

Quantile Range, robust

Polarization Measurement 3 (**weighted IQR**) for Each Industry by Time

Quantile Range weighted by transaction value, robust

Industry
- Academic & Educational Services
- Basic Materials
- Consumer Cyclicals
- Consumer Non-Cyclicals
- Energy
- Financials
- Healthcare
- Industrials
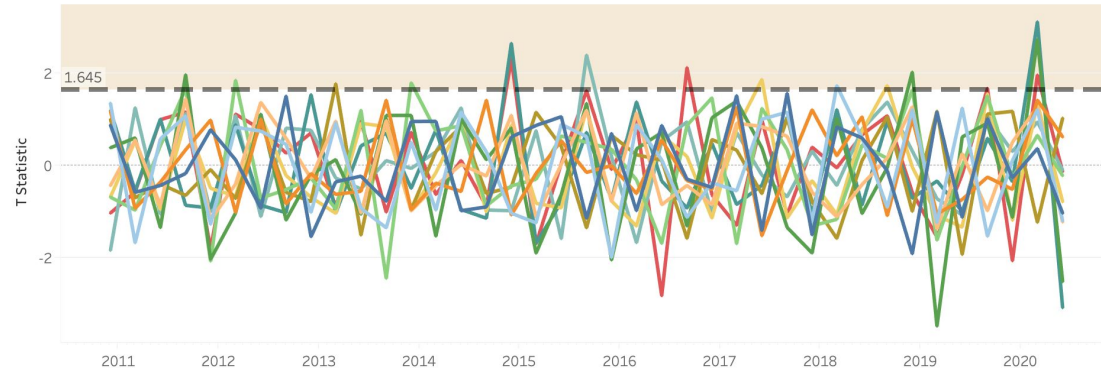- Real Estate
- Technology
- Utilities

20

3.3 Differentiate significant polarized topics from noise

- In order to identify significant polarizations, we identify significant spikes in plots of the two IQR metrics (previous slide)

- To find significant spikes, we draw a horizontal line to show the testing result (differentiate significant result from noise).

- **Significant spikes** are those that fall into the **yellow-shaded area**

- **Significant polarized topic** is represented by the **Significant spikes**

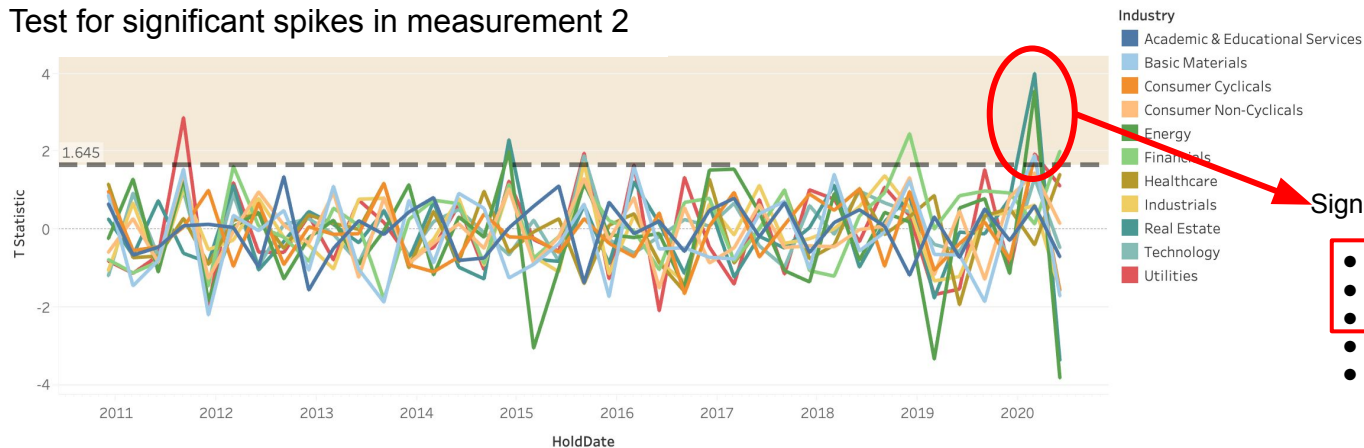Test for significant spikes in measurement 2

Test for significant spikes in measurement 3

21
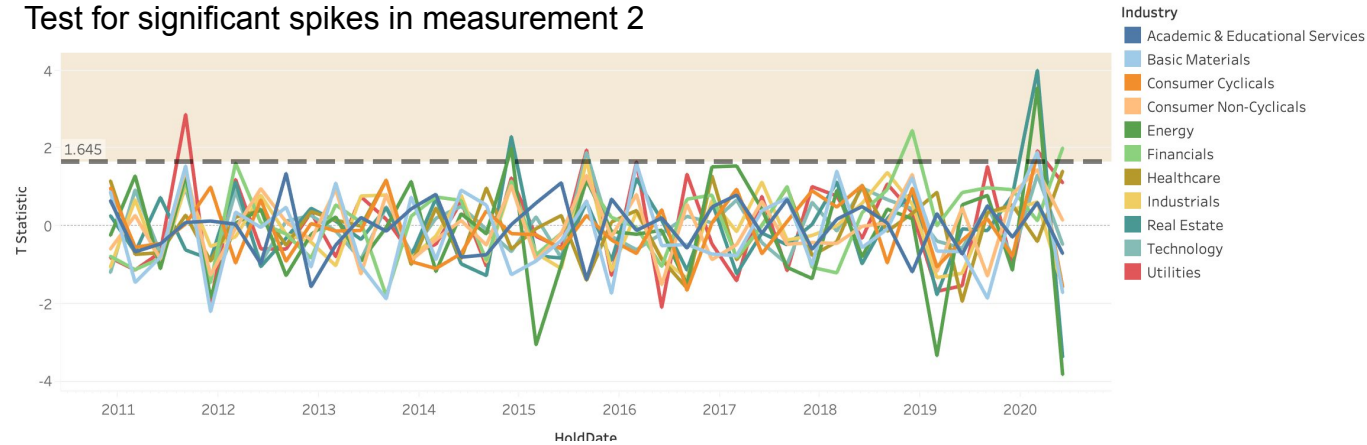
## 3.4 In 2020 Q1, Investors behave differently in real estate, energy, and utilities

Test for significant spikes in measurement 2

Industry
- Academic & Educational Services
- Basic Materials
- Consumer Cyclicals
- Consumer Non-Cyclicals
- Energy
- Financials
- Healthcare
- Industrials
- Real Estate
- Technology
- Utilities

Significant topics from measurement 2:

- **Real Estate**
- **Energy**
- **Utilities**
- Basic Materials
- Consumer Cyclicals

Test for significant spikes in measurement 3
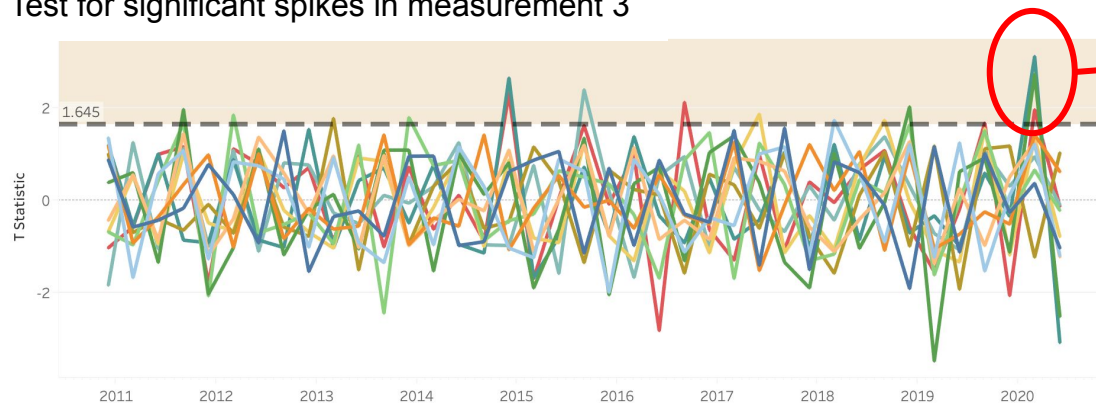
Significant topics from measurement 3:

- **Real Estate**
- **Energy**
- **Utilities**

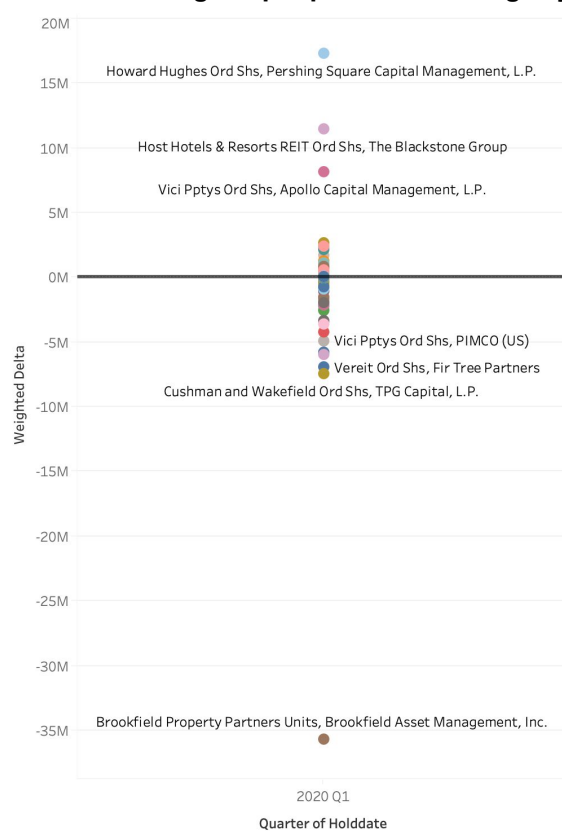# 3.5 Zoom in to the Investment in Real Estate @ 2020 Q1

Change in pct portfolio holding

Change in pct portfolio holding adjusted by actual dollar value

Each dot represents a pair of (Instrument, Investor)

04 | Polarized Topic Analysis

4.1 The characteristics of investors who drive polarized topics

👍 **Larger in**

- percentage of portfolio investment in mid-cap instruments
- concentration on top 20% of its instruments
- percentage of employees with the investment style of Emerging Market

👎 **Smaller in**

- number of positions
- percentage of employees with the investment style of Index

📌 We should pay attention to investors with the above listed features, since they reveal potential polarized topics in investment.

Example Investors Who Drive the Polarized Topic:
- Anchorage Capital Group, L.L.C.
- CarVal Investors L.P.
- Apollo Capital Management, L.P.
- Bracebridge Capital, LLC
- BlueMountain Capital Management, LLC

4.2 The investors dissimilar to Vanguard tend to drive polarized topics

- Investors in cluster B (dissimilar to Vanguard) tend to **disagree** with the market mean.
- This cluster contains 54 investors.
- Analysts should pay attention to the changes in the portfolio holdings of these investors, since they are likely to drive polarized topics.

Example Investors in Cluster B:
- Apollo Capital Management, L.P.
- King Street Capital Management, L.P.
- Angelo, Gordon & Co., L.P.
- Brookfield Asset Management, Inc.

05 | Dashboard Demo

1. **Why polarized topic is important**
   a. When they have a significant deviation than vanguard,
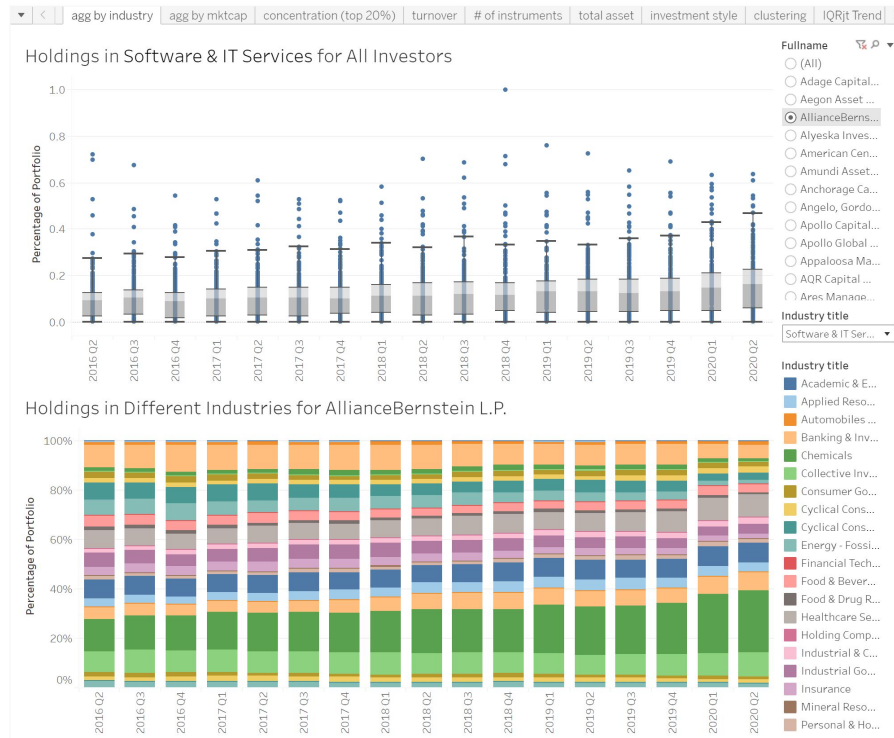   b.
2. **Add next steps**

Dashboard for Our Findings

- 7 Features
- Clustering Result
- Polarized Topics

Dashboard Demo Link

06 | **Thanks for Listening!**

Q&A

07 | Appendix

7 features of investors

The percentage of portfolio time series by aggregating instruments on industry

- Classification of Instruments (defined by the database provider)
  - 11 sectors and 30 categories

- Sample Sectors

ID	50	51	52
Sector	Energy	Basic Materials	Industrials

- Sample Categories

ID	5010	5020	5030	5110
Category	Energy - Fossil Fuels	Renewable Energy	Uranium	Chemicals

- Sample Portfolio Data

LGCYINVESTORID	FULLNAME	HOLDDATE	5010	5020	5030	5110	5120	5130	...
2004260	The Vanguard Group, Inc.	9/30/2018	0.556661	0.258986	0.001568	0.150699	0.000145	0.031941	...

7 features of investors

The percentage of portfolio in top 20% instruments time series
- Concentrated Investment
  - Many Investors tend to invest on a few instruments
  - Some Investors prefer distributed investments
  - Higher Proportion of Top 20% Instruments -> "Concentrated Investment"

- Sample Data

LGCYINVES TORID	FULLNAME	9/30/2017	12/31/2017	3/31/2018	6/30/2018	9/30/2018	12/31/2018	...
2004260	The Vanguard Group, Inc.	0.886606	0.88868	0.890548	0.886068	0.890253	0.897903	...

7 features of investors

The quarterly turnover rate time series
- Turnover Rate
  - The percentage of portfolios that have been replaced in a given period of time.
  - Higher turnover rate indicates more active investment.
  - We take one quarter as our time slot.
- Level of Turnover
  - LOW (< 50)
  - MEDIUM (50 to 100)
  - HIGH (>100)

- Sample Data

LGCYINVESTORID	FULLNAME	EffectFromDate	EffectToDate	OwnTurnover	OwnTurnoverRtng
2004260	The Vanguard Group, Inc.	7/1/2018	9/30/2018	8.1095	LOW

7 features of investors

The number of instruments time series
- Describe whether the investor held many kinds of instruments
- Sample Data

LGCYINVES TORID	FULLNAME	9/30/2017	12/31/2017	3/31/2018	6/30/2018	9/30/2018	12/31/2018	...
2004260	The Vanguard Group, Inc.	4065	4057	4091	4124	4160	4145	...

The total assets time series
- Indicate the size of the company.
- Sample Data

LGCYINVES TORID	FULLNAME	9/30/2017	12/31/2017	3/31/2018	6/30/2018	9/30/2018	12/31/2018	...
2004260	The Vanguard Group, Inc.	2.14E+12	2.29E+12	2.29E+12	2.39E+12	2.56E+12	2.22E+12	...

7 features of investors

Investment style distribution of employees

- Investment Style
- The count of employees that invest different types of investments.

- Sample Data of The Vanguard Group, Inc. (2004260)

Aggres. Gr.	Arbitrage	Convertible Arbitrage	Emerging Markets	GARP	Growth	Income	Index	Private Equity
0.085714	0.157143	0.15	0.028571	0.264286	0.007143	0.078571	0	0
Quantitative	**Sector Rotational**	**Shareholder Activist**	**Short Selling**	**Social Awareness**	**Special Situations**	**Technicals**	**Value**	**Venture Capital**
0.007143	0.007143	0.007143	0.142857	0	0	0	0.057143	0.007143