

# Data and Innovation

Jake Thornock  
BYU



# Here are my recommendations:

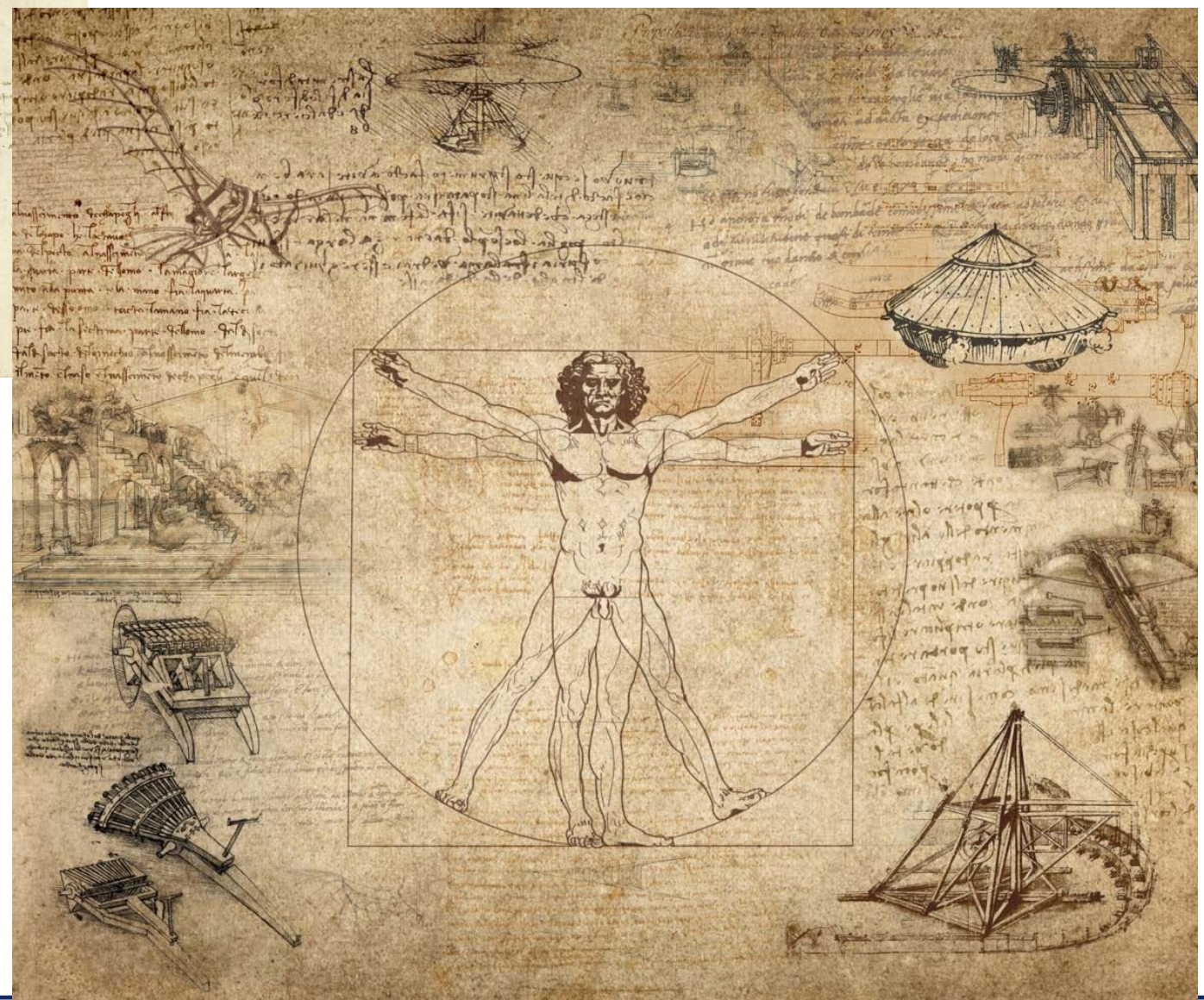
1. Develop a curious mind!
2. Always be looking for data; *anything* can be data.
3. Look for data where no one else is looking.
4. Look for new and innovative research designs!
5. Always know where the frontier is.
6. Develop skills in data analyses (see Michelle's discussion)
7. Learn the art (writing) behind the science (data analysis).



# Recommendation #1: Develop a curious mentality

- One the best parts of this job is that we get paid to scratch our curious itches.
- You are getting paid for your creativity.
- Your PhD is a doctorate in the tools of creative exploration.
- I consider myself an entrepreneur ... with excellent health insurance.







# How did Leonardo Da Vinci **create**? By letting his mind wander ... and noticing things

- *Speaking of the seemingly unrelated musings/drawings in Leonardo's notebooks, Walter Isaacson writes: "The juxtapositions can seem haphazard, and to some extent they are; we watch his mind and pen leap from an insight about mechanics, to a doodle of hair curls and water eddies, to a drawing of a face, to an ingenious contraption, to an anatomical sketch, all accompanied by mirror-script notes and musings. But **the joy of these juxtapositions is that they allow us to marvel at the beauty of a universal mind as it wanders exuberantly in free-range fashion over the arts and sciences and, by doing so, senses the connections in our cosmos. We can extract from his pages, as he did from nature's, the patterns that underlie things that at first appear disconnected.**"*

# Generating ideas

- Read outside of accounting, apply it to accounting.

Don't sweat ideas. If you're active and curious, *they will come!*

- Right now:  $\text{TIME} > \text{IDEAS}$
- Soon:  $\text{IDEAS} > \text{TIME}$ .



# Generating ideas

- Give yourself time to think!
  - Aswath Damodaran: “I have discovered that my mind is most productive when I am not trying too hard to be insightful.”
  - Leonardo: “Men of lofty genius *sometimes accomplish the most when they work least* for their minds are occupied with their ideas and the perfection of their conceptions, to which they afterwards give form.”
- Learn to fluidly move between the macro and the micro.

## Recommendation #2: Always be on the lookout for data; data is everywhere

- As a PhD student, I found a working paper on international earnings announcements
  - Landsman, Maydew and Thornock (2012 *JAЕ*)
- As a PhD student, I discovered that Thompson Reuters had *foreign institutional ownership*
  - We wrote a paper on dividend withholding – it died
  - But it helped us discover foreign flows data, which led to
    - Hanlon, Maydew and Thornock (2015 *JF*)



# Always be on the lookout for data

- As a PhD student, a finance professor told me about Google Trends.
  - Drake, Roulstone and Thornock (2012 *JAR*)
  - Drake, Jennings, Roulstone and Thornock (2017 MS)
- From that paper, we asked the SEC for EDGAR search data
  - Drake, Roulstone and Thornock (2015 CAR)
  - Drake, Roulstone and Thornock (2016 *JAЕ*)
- Those papers led to a consulting project, from whence I discovered IP address ownership data
  - Bozanic, Hoopes, Thornock and Williams (2016 *JAR*)

# You can even get ideas from watching TV



# March Market Madness: The Impact of Value-Irrelevant Events on the Market Pricing of Earnings News\*

MICHAEL S. DRAKE, *Brigham Young University*

KURT H. GEE, *Stanford University*

JACOB R. THORNOCK, *University of Washington*

## ABSTRACT

Each year, the NCAA basketball tournament (March Madness) is a daytime distraction for millions of people, providing a largely exogenous shock to investor attention. We investigate whether March Madness influences the market response to earnings by diverting investor attention away from earnings news. We find that the price reaction to earnings news released during March Madness is muted. This result generally holds across several samples and additional analyses. We also find that the result is more muted for low institutional ownership firms, consistent with the effect being driven by less-sophisticated investors. Furthermore, we find that it takes the market 30 to 60 days to correct for the distraction effect. Overall, we provide a unique test of the theory of limited attention by documenting that extraneous events can have a significant impact on the pricing of earnings.

*Contemporary Accounting Research* Vol. 33 No. 1 (Spring 2016) pp. 172–203 © CAAA  
doi:10.1111/1911-3846.12149



## Recommendation #3:

Look for data where no one else is looking.

*So, where can I find new data?*

- **NOT** WRDS
- There are fewer and fewer meaningful papers that to be written using
  - ETRs
  - BTDS
  - UTBs



# Product differentiation

- You're an entrepreneur. Papers (and their data) are your *product*. How do you differentiate them?
- Or maybe you're an arbitrageur. How do make money? => *Go where the excess returns live.*
  - For every paper, ask: how will this generate *alpha*?

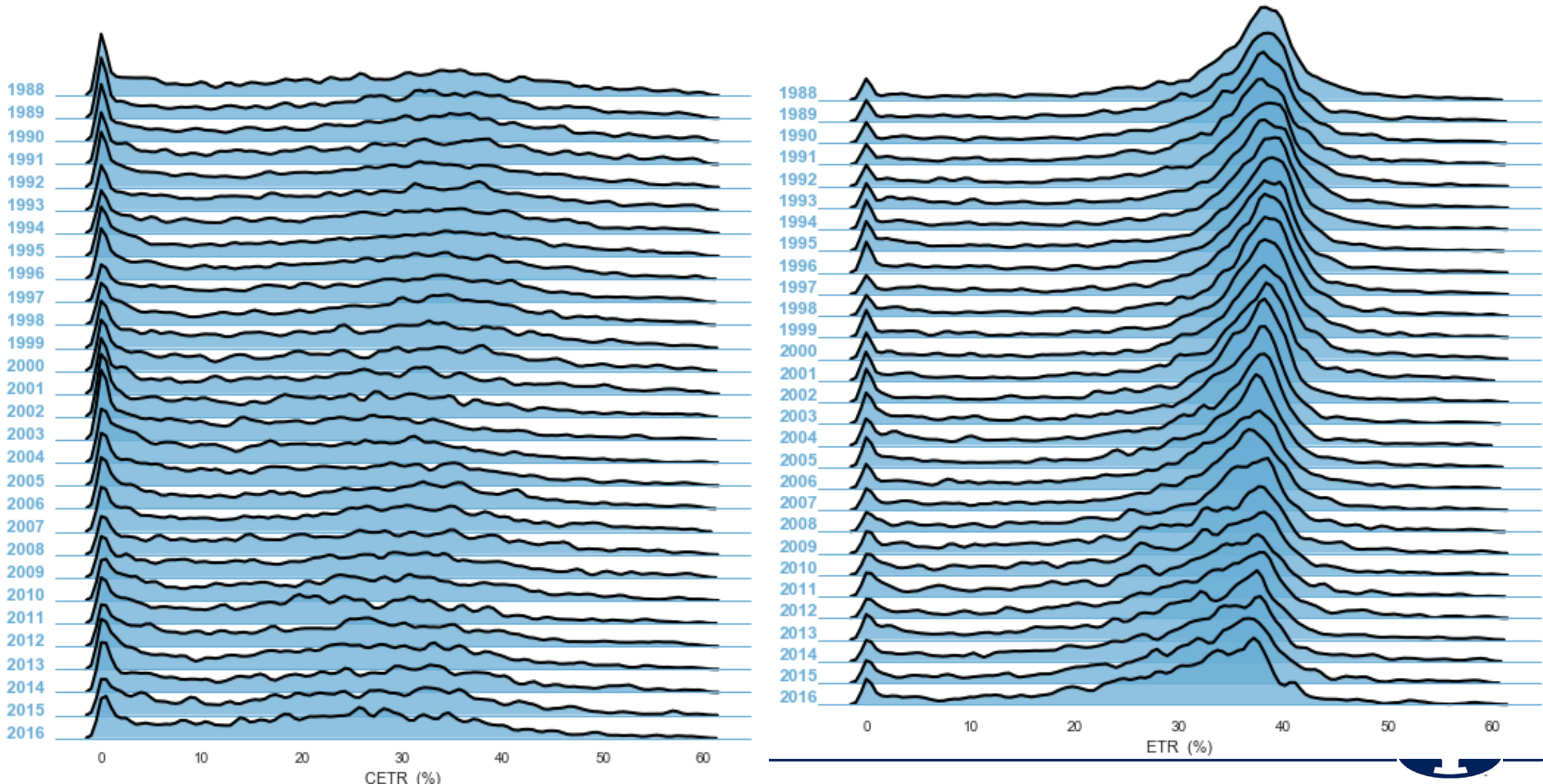
# Product differentiation

- Peter Theil: 0 to 1 .... Or .... 1 vs  $n$
- “You don’t want to be the fourth online pet food company, you do not want to be the twelfth solar panel company and you do not want to be the  $n$ th company of any particular trend.”
- “What I prefer over trends is a sense of mission. That you are working on a unique problem that people are not solving elsewhere. Unique missions are much preferred over trends”

# Where to find new data?

- “Anecdata” can often lead to actual data
- Look at *market controversies* in the news
  - E.g., Dark trading (Jing Pan, University of Utah)
- Look at accounting/tax controversies from practice
  - There will, no doubt, be a number of papers on this new tax law change.
- Look at economics blogs
- Twitter
  - *#dataiscool*

# Dyreng and Thornock 2018





In 30 minutes of Googling, I found the following data sources:

What emails were being sent before the discovery of fraud at Enron?

- <https://aws.amazon.com/datasets/enron-email-data/>

What do people look at on Wikipedia?

- <https://aws.amazon.com/datasets/wikipedia-traffic-statistics-v2/>.

Employee hiring/firing/pension data

- [https://catalog.data.gov/dataset/ebsa-form-5500-annual-return-report-filing-enforcement-data#topic=finance\\_navigation](https://catalog.data.gov/dataset/ebsa-form-5500-annual-return-report-filing-enforcement-data#topic=finance_navigation)



- What leads to bank failure?
  - <https://catalog.data.gov/dataset/fdic-failed-bank-list>
- Pension insurance data, including data on claims, locations, employees
  - <https://www.pbgc.gov/prac/data-books>
- Lawsuits initiated by the SEC [only 2010]
  - <https://catalog.data.gov/dataset/2010-sec-litigation-releases>
- FED data on just about anything macro
  - <https://www.federalreserve.gov/datadownload/help/default.htm>



- Searchable NYTimes articles, going back to 1851.
  - <http://developer.nytimes.com/>

## Recommendation #4: Look for new and innovative research designs!

- Old data, new research design
- Famous person: "There are rarely *new questions*; there are more often **new tests** and **new data** to answer old questions."



# Cai Szeidl QJE 2018

We organized business associations for the owner-managers of young Chinese firms to study the effect of business networks on firm performance. We randomized 2,820 firms into small groups whose managers held monthly meetings for one year, and into a “no-meetings” control group. We find that: (1) The meetings increased firm revenue by 8.1 percent, and also significantly increased profit, factors, inputs, the number of partners, borrowing, and a management score; (2) These effects persisted one year after the conclusion of the meetings; and (3) Firms randomized to have better peers exhibited higher growth. We exploit additional interventions to document concrete channels. (4) Managers shared exogenous business-relevant information, particularly when they were not competitors, showing that the meetings facilitated learning from peers. (5) Managers created more business partnerships in the regular than in other one-time meetings, showing that the meetings improved supplier-client matching.



# Lawrence Ryans Sun Laptev 2018

- This study presents a **field experiment** we conducted in which media articles *for a random sample of firms with earnings announcements are promoted to a one percent subset of Yahoo Finance users.*
- ...we find that promoted firms have higher abnormal returns on the day of the earnings announcement, and some evidence of lower bid-ask spreads. Moreover, these results are more pronounced for less visible firms, negative earnings news, and on days with fewer promoted firms.
- These findings suggest that investor attention affects the pricing of earnings and that retail investors buy stocks that catch their attention, in a setting **where attention is randomly assigned.**

# Recommendation #5:

## Always know where the frontier is.

- Your “new” data/ideas might not be that new!
- Marc Andreessen (founder Netscape)
  - Q: Do frontiers still exist? “Yes, but these things look like cults and fringe activities until they break into the mainstream.”
  - It used to be that frontiers were geographic. “...instead what we have are virtual frontiers, intellectual frontiers, creative frontiers, entrepreneurial frontiers, technological frontiers...”

# Be original

- Lack of contribution is usually the death blow of a paper.
- Mediocrity is the career killer.
- Ask yourself and *be brutally honest*: If you stumbled across your own article, would YOU want to read it?



# Recommendation #6:

## Develop skills in data analyses (see Michelle's discussion)

- The data and research skills of 20 years ago are dying (at least for for empirical/archival research)
- Face it: The data and research skills of today have a limited shelf life
- Face it: the bar will move!
- Where do I see us going? *I don't know*, but I'm betting on analytics
  - Python
  - R

## Recommendation #7:

Learn the art (writing) behind the science (data analysis).

- Good data + bad writing = rejection
- “There is a simple way to package information that, under the right circumstances, can make it irresistible. All you have to do is find it.” Malcolm Gladwell, *Tipping Point*
- Good artists don’t learn to draw, they learn to see. Ed Catmull (Co-founder, Pixar)

# Writing tips

- Assume your reader is super smart ... but knows nothing.
- Spell out your implications ... don't make the reader.
- Start with the priors of the *reader*, not the authors
- Accessible writing > big, dense words
- Master the short, crisp sentence.



# Writing tips

- Cultivate story-telling
- Netflix vs textbooks – the former is popular, the latter are not.
- People want stories! Even if your subject matter is challenging
  - Michael Lewis
  - Neal Degrasse Tyson
  - William McCullough



# To recap: Here are my recommendations:

1. Develop a curious mind!
2. Always be looking for data; *anything* can be data.
3. Look for data where no one else is looking.
4. Look for new and innovative research designs!
5. Always know where the frontier is.
6. Develop skills in data analyses (see Michelle's discussion)
7. Learn the art (writing) behind the science (data analysis).





# Motivational speech!!

- This job is the best!!
- Fulfillment → happiness.
- This job is truly fulfilling because
  - You get paid to be creative
  - You work with amazing and genuine people
  - You get to mentor and be mentored
  - (Teaching is pretty fun too!)





**I ILLINOIS**  
Gies College of Business

# Textual Analysis

ATA Doctoral Consortium 2018  
Michelle Hutchens

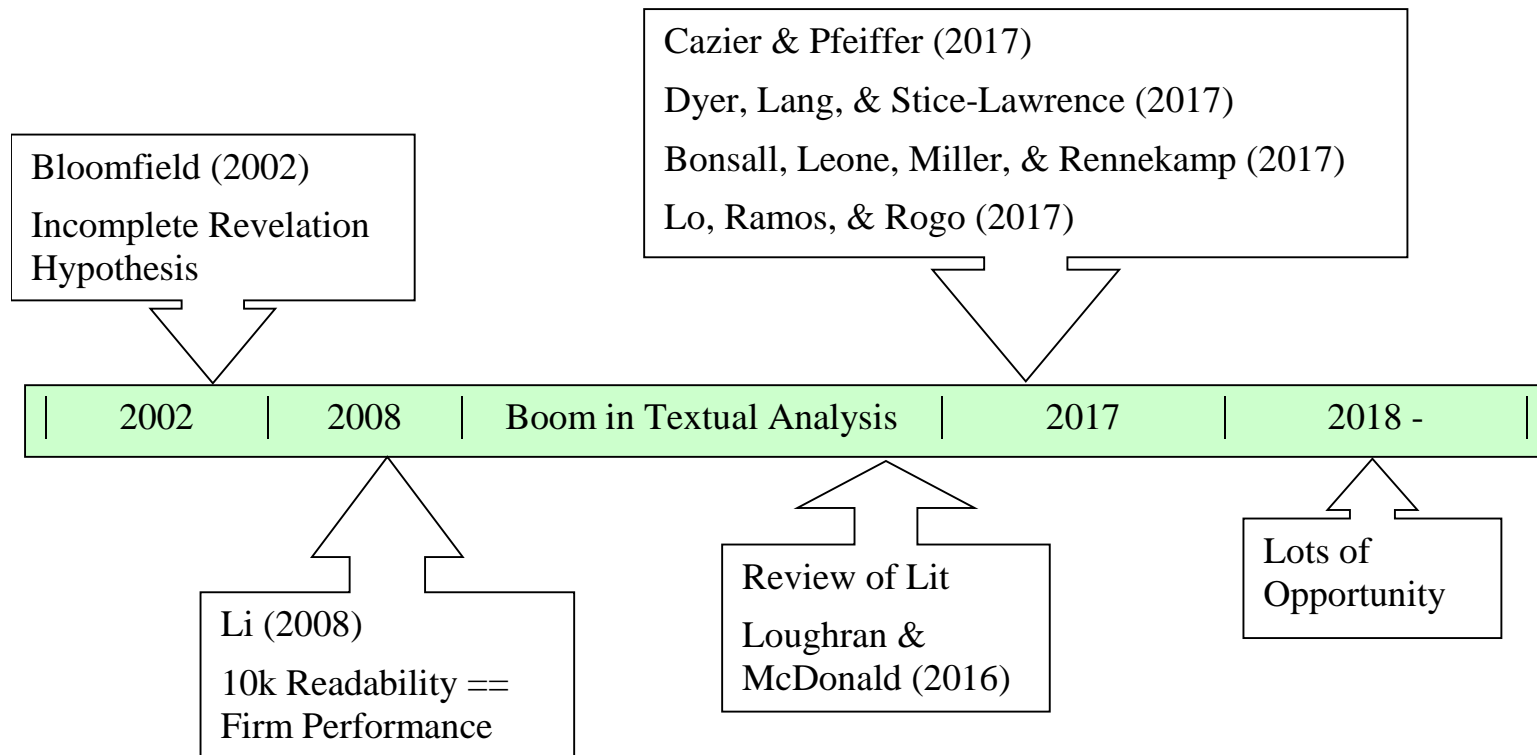
# Overview

1. Textual Analysis in the Financial Accounting Literature
2. Textual Analysis in the Tax Literature
3. Tools and Resources
4. Questions

# Textual Analysis: Financial Accounting

Texts Analyzed	Measures Extracted
10-k	FOG
MD&A	BOG
Conference Calls	Length
Press Releases	File Size
....	#s / Tables
	Tone
	Topics (LDA)
	Stickiness
	Repetitive
	...

# A Brief Timeline





# The Criticisms & Suggestions

## **Bloomfield (2008) Discussion of Li (2008)**

1. Potential Explanations
  - a. Obfuscation
  - b. Losses are more difficult to describe
  - c. Attribution to external events
  - d. Misdirection/distraction
  - e. Management by exception
  - f. Conservatism
  - g. Litigation Risk
2. Future research should examine the “why”

# The Criticisms & Suggestions

## **Textual Analysis in Accounting and Finance: A Survey Loughran & McDonald (2016)**

1. Motivate research by hypotheses “closely tied to economic theories”
2. Define your construct of interest. What is “readability”?  
*Note: The traditional measures of readability in linguistics is not a distinguishing characteristic in interpretation of financial texts.*
3. The Business vs. The Document
4. Minimize “econometric exorcism” and clearly articulate data cleansing steps

# The Criticisms & Suggestions

## Dyer, Lang, & Stice-Lawrence (2017) Discussion of Cazier & Pfeiffer (2017)

1. Need to focus on causality
2. Multiple methodology papers (e.g., Bonsall et. al 2017)
3. Lack of integration within existing literature
4. Suggest the following 2x2 for characterizing Textual Analysis Research

Discretionary * Helpful	Non-Discretionary * Helpful
Discretionary * Harmful	Non-Discretionary * Harmful

# The Criticisms & Suggestions

## Miller (2017) Discussion of Dyer, Lang, & Stice-Lawrence (2017)

1. “I believe the evidence presented suggests the *need for further research* in the area, but I will argue that it would be premature to provide policy suggestions based on this evidence.”
2. Concerns regarding author discretion in LDA (intentional and unintentional)
3. Textual attributes may be harmful to some users while being helpful to others
4. Are we studying the textual attributes that practitioners care about?

# Textual Analysis: Tax Literature

Texts Analyzed	Measures Extracted
10-k	FOG
MD&A	BOG
Conference Calls	Length
Press Releases	File Size
....	#s / Tables
Tax Footnote	Tone
	Topics (LDA)
	Stickiness
	Repetitive
	...

Tax Topic Items, etc.



# Tools and Resources

## Step 1: Learning PERL/PYTHON/other



– 183 Python courses from UIUC, Michigan, UW, etc.



or google it ☺

[https://rstudio-pubs-static.s3.amazonaws.com/79360\\_850b2a69980c4488b1db95987a24867a.html](https://rstudio-pubs-static.s3.amazonaws.com/79360_850b2a69980c4488b1db95987a24867a.html)

## Step 2: Downloading Texts to be Analyzed

- PERL/PYTHON to download data
- Sample code available (google it ☺)

-<http://www.wrds.us/index.php/tutorial/view/26,%20http://www.kaikaichen.com/?p=59>

## Step 3: Clean/Cut/Sort Text/Extract Measures of Interest

- Be mindful of intentional and unintentional choices that may impact your analysis
- Document, so others can replicate
  - See Appendix C in Bonsall et. al (2017)



# Questions ?

**I** ILLINOIS  
Gies College of Business





# Thank You

**I** ILLINOIS  
Gies College of Business