

# Data Mining and Text Mining

Rei Sanchez-Arias, Ph.D.

[rsanchezarias@floridapoly.edu](mailto:rsanchezarias@floridapoly.edu)



**Adam Rothman**  
@arothmanhistory

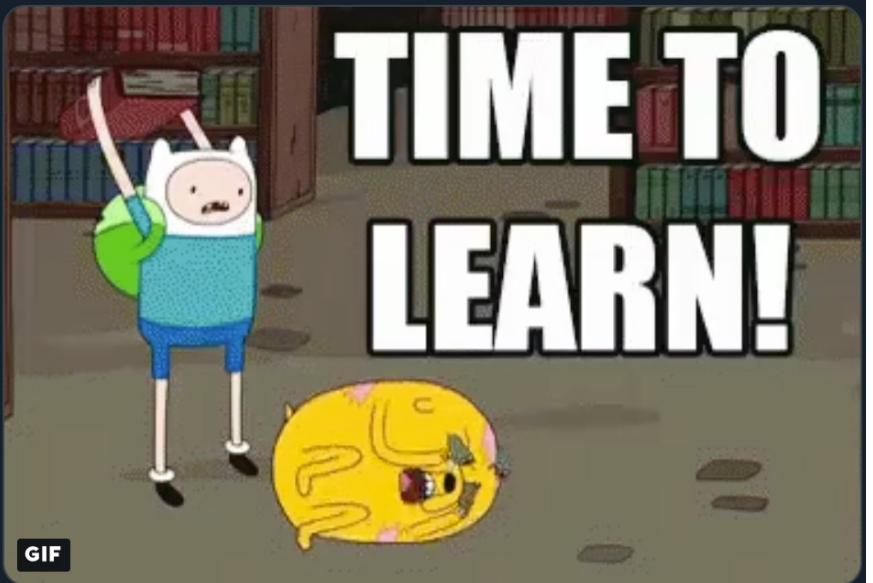
Don't think of it as the first day of class. Think of it as a syllabus reveal party!

5:15 PM · Jan 7, 2019 · Twitter for Android



**Yoursteppops** 🇺🇸 @PushaTeeShirt · 2h

Every MSU professor skipping **syllabus** right to chapter 1. 😂



**chad** @bryanwe · 3h

My entomology professor started her **first lecture** with a slide that said "Things that bug me" this is gonna be a good semester

1 8 1 1



**daddy** @dadd\_ilf · Jan 7

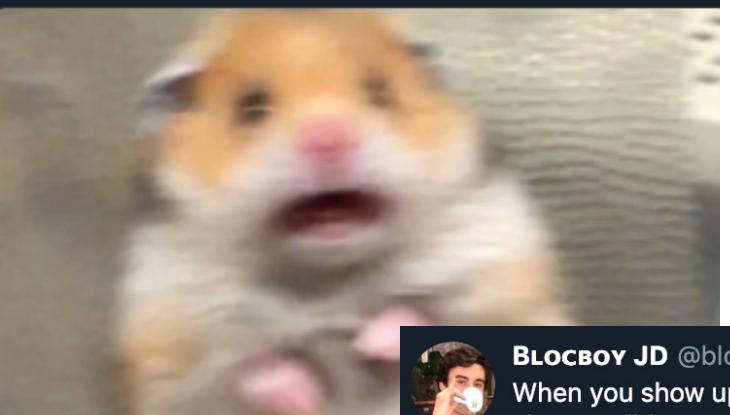
in my **first lecture** my professor literally said "if u just wanna do the polls from your room and not come to class that's fine i guess" ?? sir don't encourage me

1 17 1



**LN** @Ellenweaver14 · 11h

When the **professor** finishes going over the **syllabus** and instead of ending class an hour early, starts actual lecture that requires note taking



**Blocboy JD** @blocboijjd · 1h

When you show up to class on **syllabus** week and the **professor** pulls up chapter slides



**sarah** @CrangH · Dec 7

**Professor**, how was your weekend?

"It's on the **syllabus**."

1 8 1 1



**ash** 📱 @makishley · 17h

why is my **professor** using a whole hour and fifteen minute class to go over a 5 page **syllabus**



1 12 1 1



**virgin** @lcantdrive · 23h

Calculus **professor** started the **first lecture** on the **first day of class**.  
Electric chair!!!



**Kassari** @Inshallah · 11h  
Me: \*walks into class on **syllabus** day\*

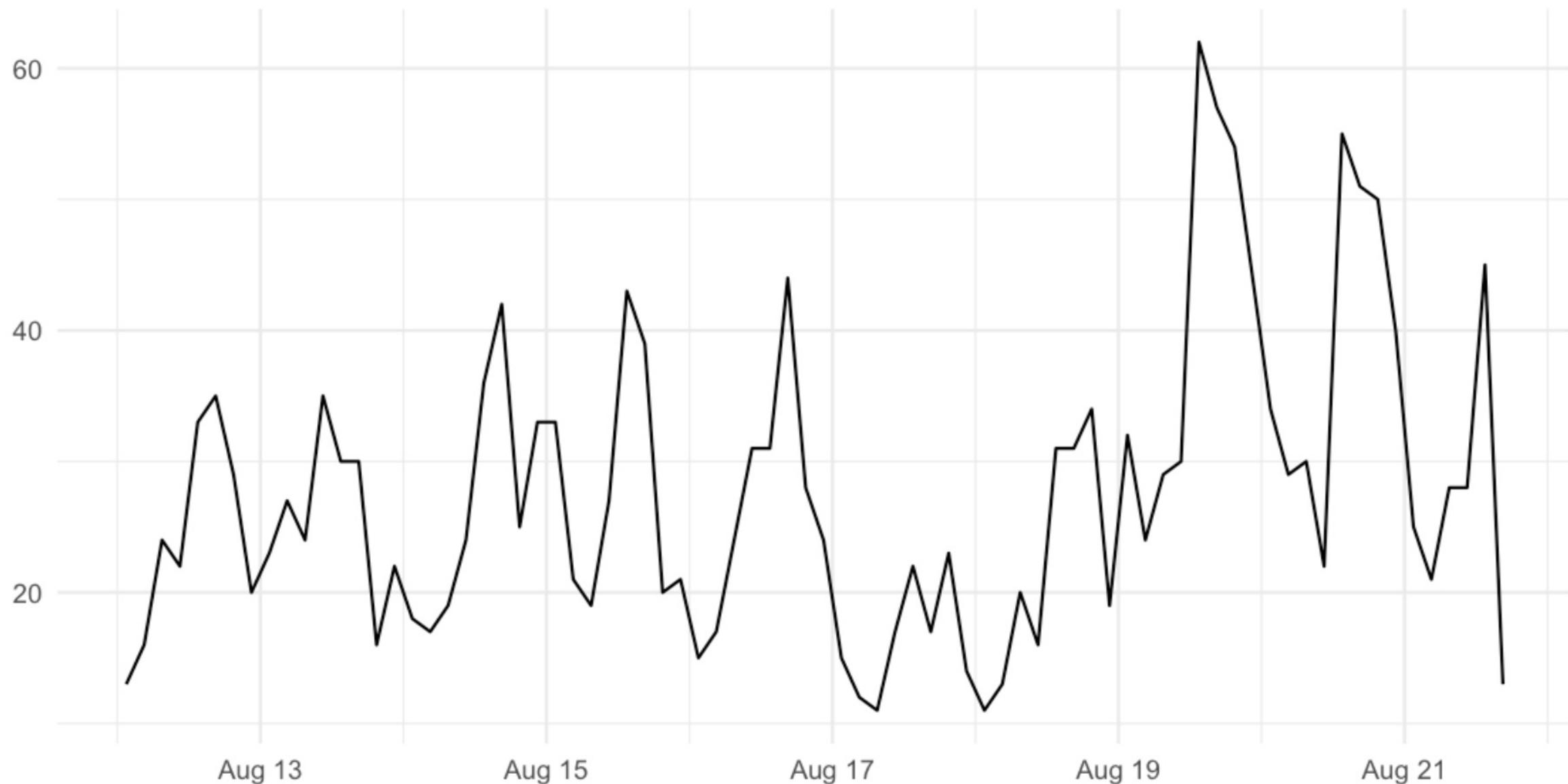
Professor: ok students, I already uploaded the **syllabus** online so turn to chapter 3 in your textbook

Me:



# Frequency of "first lecture" Twitter statuses from past 9 days

Twitter status (tweet) counts aggregated using three-hour intervals



Source: Data collected from Twitter's REST API via rtweet

# Tweeting about syllabus in the US

```
syllabus_tw %>%  
  filter(retweet_count > 50, is_retweet == FALSE)
```

text  
<chr>

Academics: It's not too late to check your syllabus and make sure there are works by women and people of color scholars on there. So, go ahead and double-check and don't give me any excuses for why they aren't there. You can use Google scholar just as well as I can.

When you jump right into lectures during syllabus week <https://t.co/ABcHCJOo2p>

Teaching computational ethics, data science, data journalism, or technology & society this semester? May I recommend a book for your syllabus? <https://t.co/KleFoDib32>

when my professor starts teaching me something other than the syllabus on day one <https://t.co/pLSmbJklf8>

@caromholland @rsl318 And dear grad seminar faculty: perhaps the kindest practice I've ever encountered is to require anyone who wants to talk on content not on the syllabus to give everyone else some brief background on the person/concept/context. Funny how it cuts down on name-dropping...

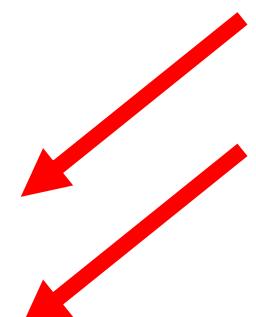
Today is a good day to retweet this, thanks to @NativeApprops and @PepePierce for their work: <https://t.co/h9pRKOfWoM>

A plea to everyone out there finishing up the syllabus for your fall computing-related class: put some ethics on it. This is a thread to tell you why you should do that and to offer some tips on how. Massive cultural change within computing starts with a single syllabus!

excerpt from a class syllabus ..... I'm ready <https://t.co/M7Vhl4sEqp>

1. Read the syllabus. Really. All kinds of important info there\n\n2. Don't be scared of office hours. They're there for you!\n\n3. Take at least 1 class just cuz it sounds fun. If not your 1st semester then soon. Because learning should be fun sometimes, & college is for exploring <https://t.co/YxI4bX0WRQ>

The most popular school-specific advice has been:\n1. Go to every class, even when you don't feel like it.\n2. Read your syllabus.\n3. Professors want you to succeed, talk to them, ask questions, go to their office hours.\n4. Do all the readings.\n5. Prepare and plan.



Collected: 2019/08/21

# Data Mining and Text Mining

Welcome to **Data Mining and Text Mining!**

Class meetings: will run on Tuesday and Thursday from 5:30 PM to 6:45 PM EST



# Textbook and Required Materials

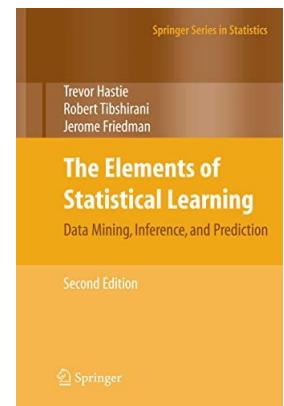
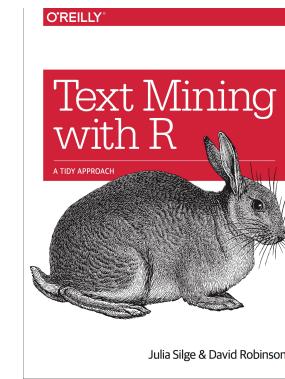
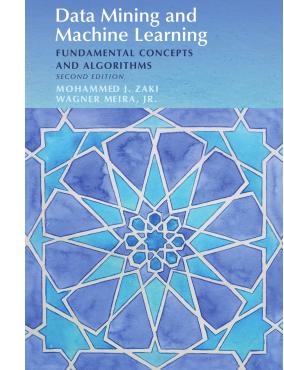
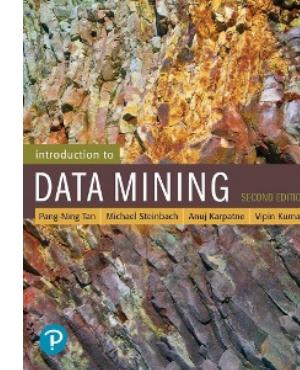


*“Text Mining with R: A Tidy Approach”* by Julia Silge and David Robinson. <https://www.tidytextrmining.com/>

## Other supplementary books (not required)

Some of the materials presented by your instructor are inspired by some of the contents in the textbooks listed below

1. "[The Elements of Statistical Learning](#): Data Mining, Inference, and Prediction" by Hastie, Tibshirani, Friedman. Second Edition (2009)
2. "[Data Mining and Machine Learning: Concepts and Algorithms](#)" by Zaki and Meira
3. "Data Mining for Business Analytics: Concepts, Techniques, and Applications" by Shmueli, Bruce, Yahav, Patel, Lichtendahl. First Edition (2018).
4. "Applied Predictive Modeling" by Max Kuhn and Kjell Johnson. Second Edition (2018)



# Equipment and Materials

This course is an *application-driven introduction to data mining and text mining*. We will use the R programming language and RStudio. **Both of these are free**. The course covers fundamental and popular R packages for data mining and text mining, introduced as working examples.

- Recommended: a local installation of can be completed by downloading R from the [R Project web site](#). The free and open-source Integrated Development Environment (IDE) for R can be downloaded [from the RStudio web site](#). You are welcome to use a different IDE but using RStudio is highly encouraged.
- Suggested: You can also create an account in <https://rstudio.cloud> to access a cloud-based version of RStudio to work in your assignments/projects. Notice that there is a limited number of hours you can use this platform every month.

**CHECK CANVAS MODULE FOR MORE DIRECTIONS**

