



MOOC Insights through Scraping edX

Patrick Masi-Phelps
August 2017

1. Massive Open Online Course (MOOC) space overview
2. Objectives
3. edX Courses and Subjects
4. Intra-Subject Distributions
5. Non-Profit University vs. Private Company vs. Non-Profit Institution/Other Organization Characteristics
6. Next Steps

Top MOOC Providers

58 million global users

The Coursera logo, featuring the word "coursera" in a blue, sans-serif font with a stylized infinity symbol or "C" shape at the beginning.

23 million



10 million



6 million

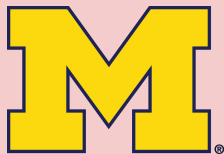


5.3 million

The Udemy logo, featuring the word "udemy" in a green, lowercase, sans-serif font.

4 million

Public and Non-Profit Universities



Private Companies



Non-Profit Companies and Other Organizations



Objectives

Understand high level characteristics of the MOOC space

- Providers - who offers MOOCs?
- Subjects - which subjects are most popular?
- Courses - what does a typical edX course look like?

Trends - how do these characteristics relate and change over time?

Who cares?

- Academic institutions
- Students



Dog Origins

Learn about the origin of dogs and how certain wolves made a journey from vermin to beloved pet.



**Not Currently
Available**

About this course

2 Reviews 4.5/5 ★★★★★

The origin of dogs is shrouded in mystery. In this course, you will learn how to interpret the often contradictory ideas about how, why, where and when dogs arose.

Geneticists, archaeologists, anthropologists and behavioral scientists offer a variety of

[See more](#)

What you'll learn

- The difference between domestication and taming
- The importance of critical periods for social imprinting
- Behavioral development
- Hunting vs. trash scavenging as the mechanism of domestication
- Belyaev's foxes

Meet the instructor



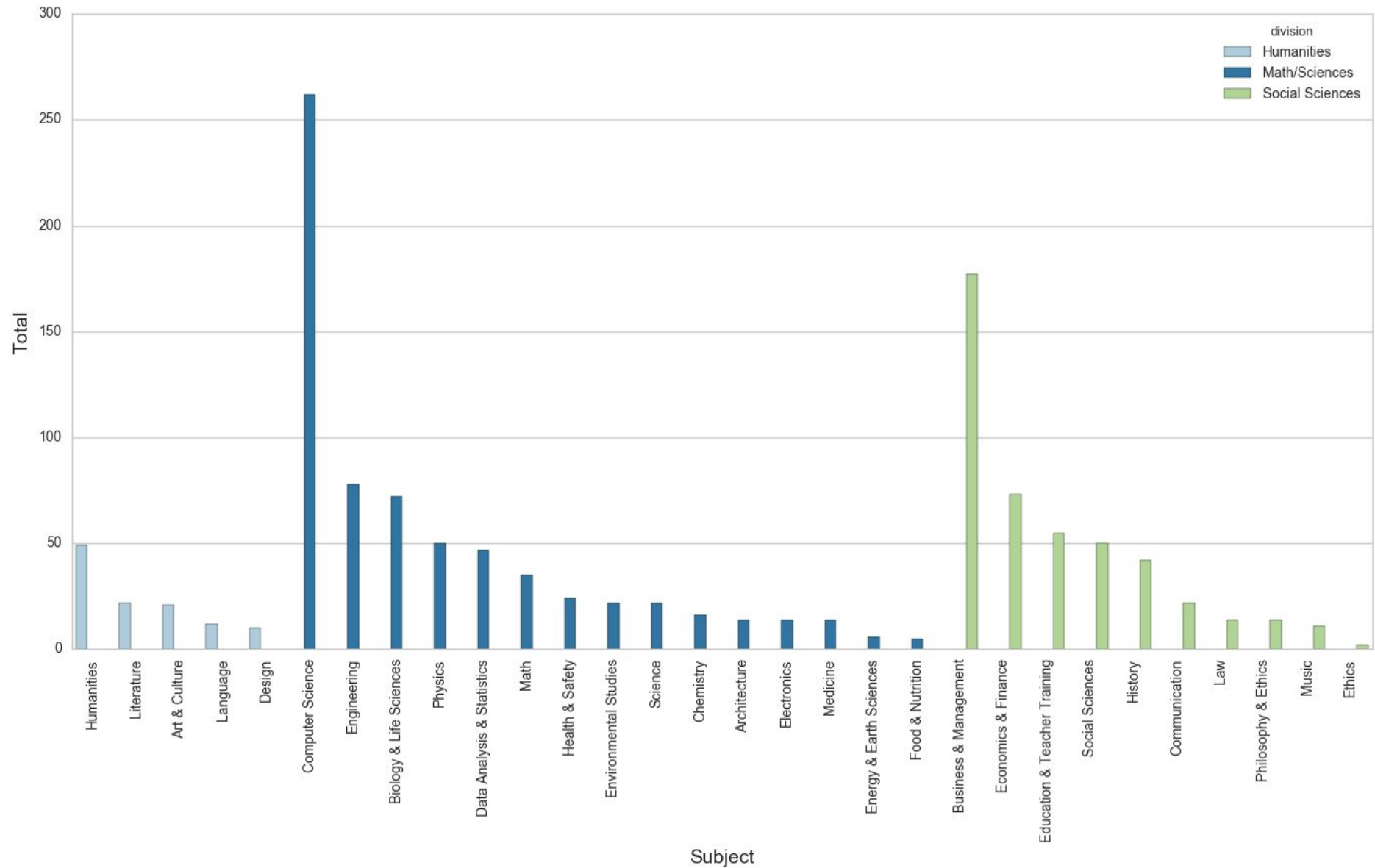
Clive D.L. Wynne, PhD
Professor of Psychology
Arizona State University

| | |
|----------------|---|
| 🕒 Length: | 3 weeks |
| 👤 Effort: | 2 - 3 hours/week |
| 📚 Type: | XSeries, Professional Education, Self-Paced |
| 💰 Price: | \$349 |
| 🏛️ Institution | ASUx |
| 🎓 Subject: | Biology & Life Sciences |
| ● Level: | Intermediate |
| 💬 Language | English |
| 📺 Video | English |
| 📄 Transcript | |

Courses

What's available?

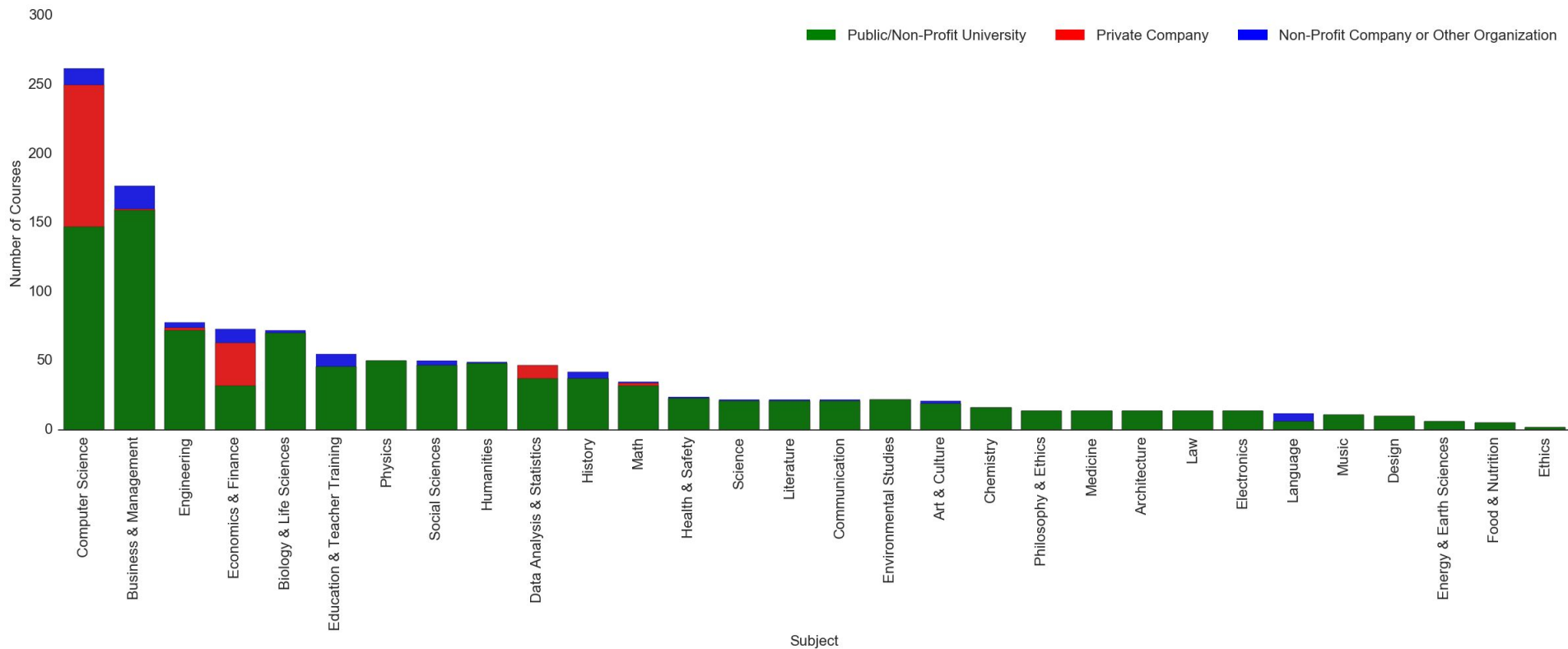
edX Courses by Subject



Providers

Which institutions provide courses in each subject?

edX Courses by Subject and Institution Type



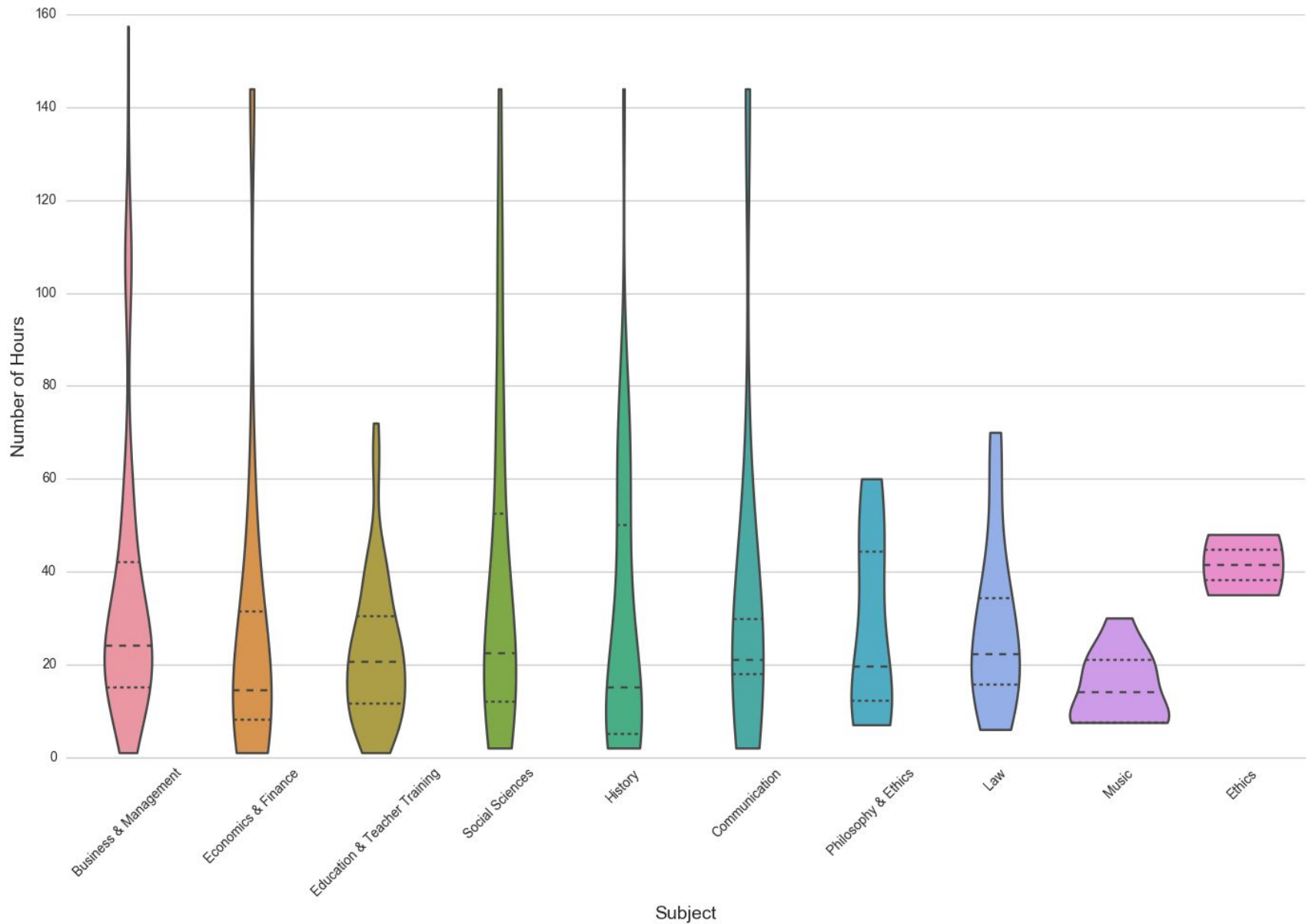
Course Characteristics

How long is a typical edX course?

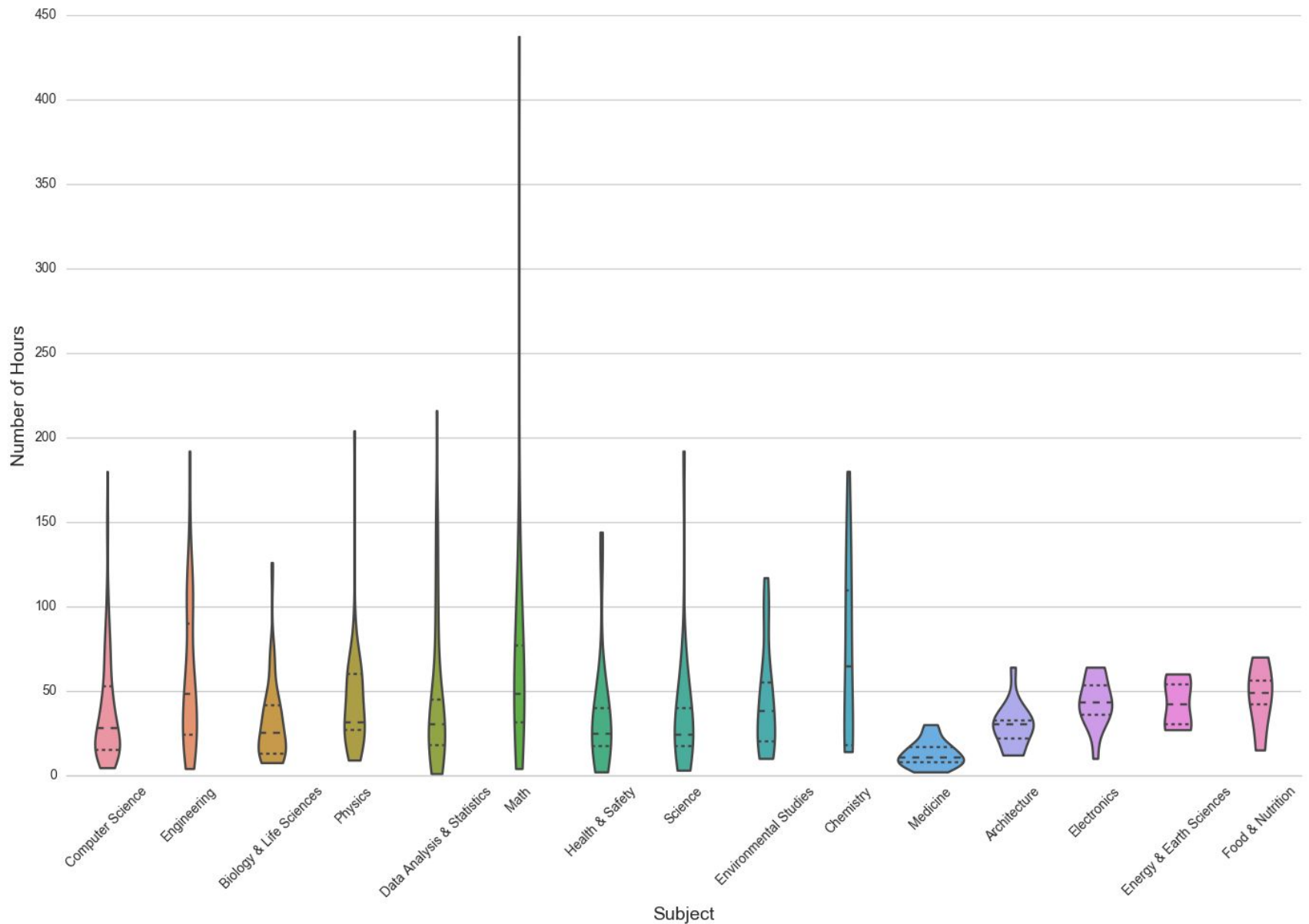
(suggested hours per week * suggested weeks to complete)

How does course length differ by subject? By provider?

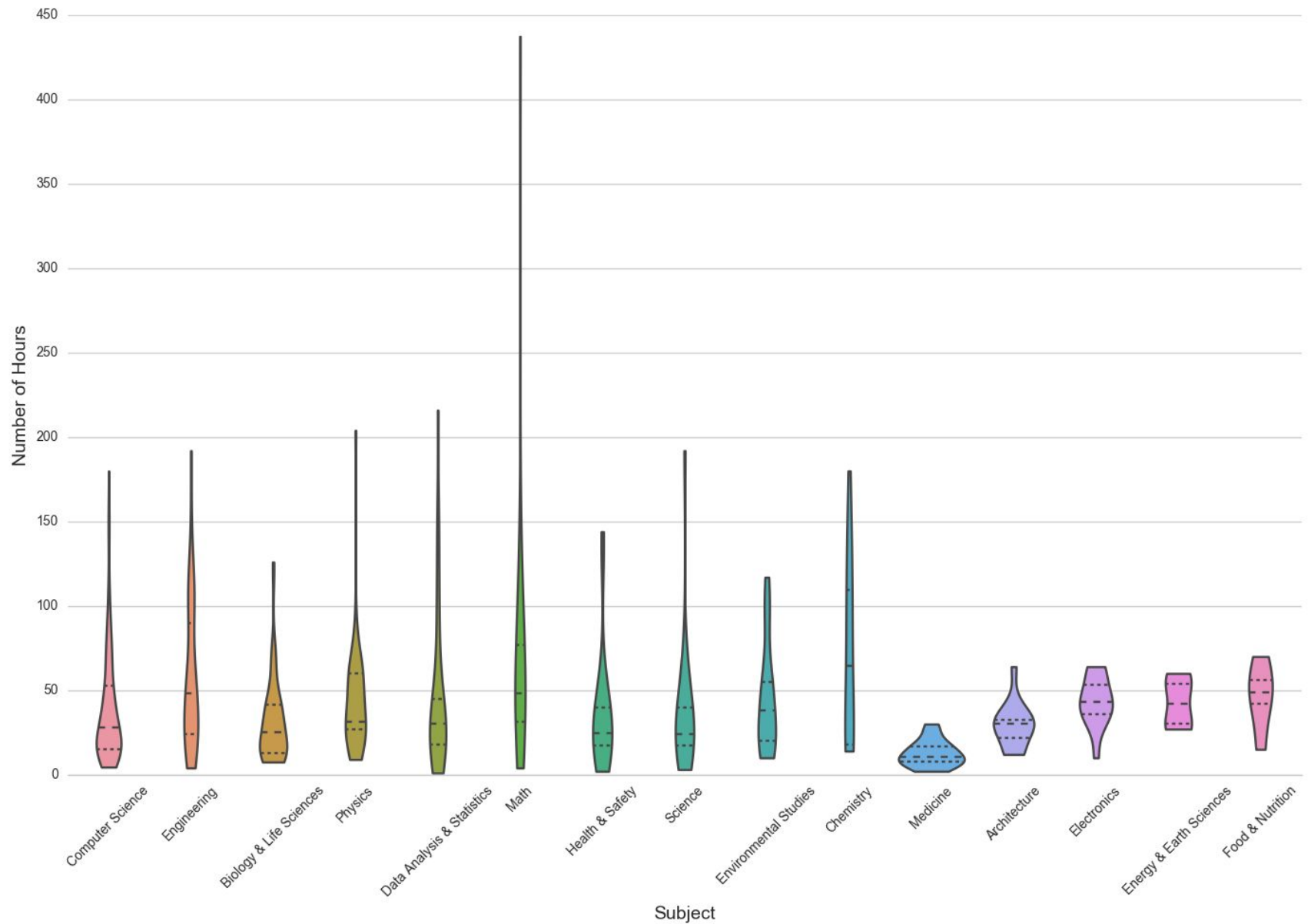
Course Hours Distribution - Social Sciences



Course Hours Distribution - Math/Natural Sciences



Course Hours Distribution - Humanities



Course Length by Provider Type

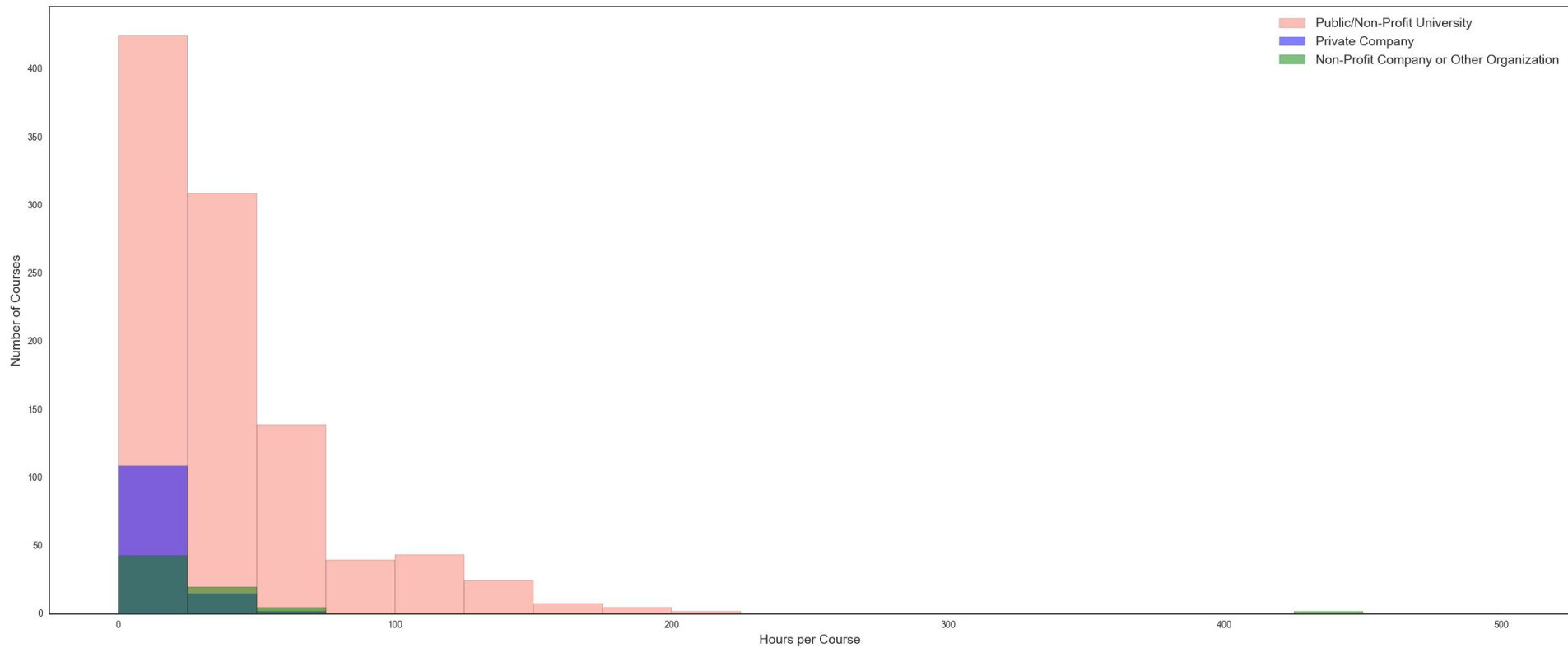
How does the length of an edX course relate to the type of provider?

Institution Types - Course Length Summary Statistics

| Type of Institution | Number of Courses Offered | Course Length (mean) | Course Length (var) |
|--|---------------------------|----------------------|---------------------|
| Non-Profit Company or Other Organization | 70 | 33.57 | 5132.77 |
| Private Company | 126 | 15.93 | 128.84 |
| Public or Non-Profit University | 997 | 40.40 | 1176.68 |

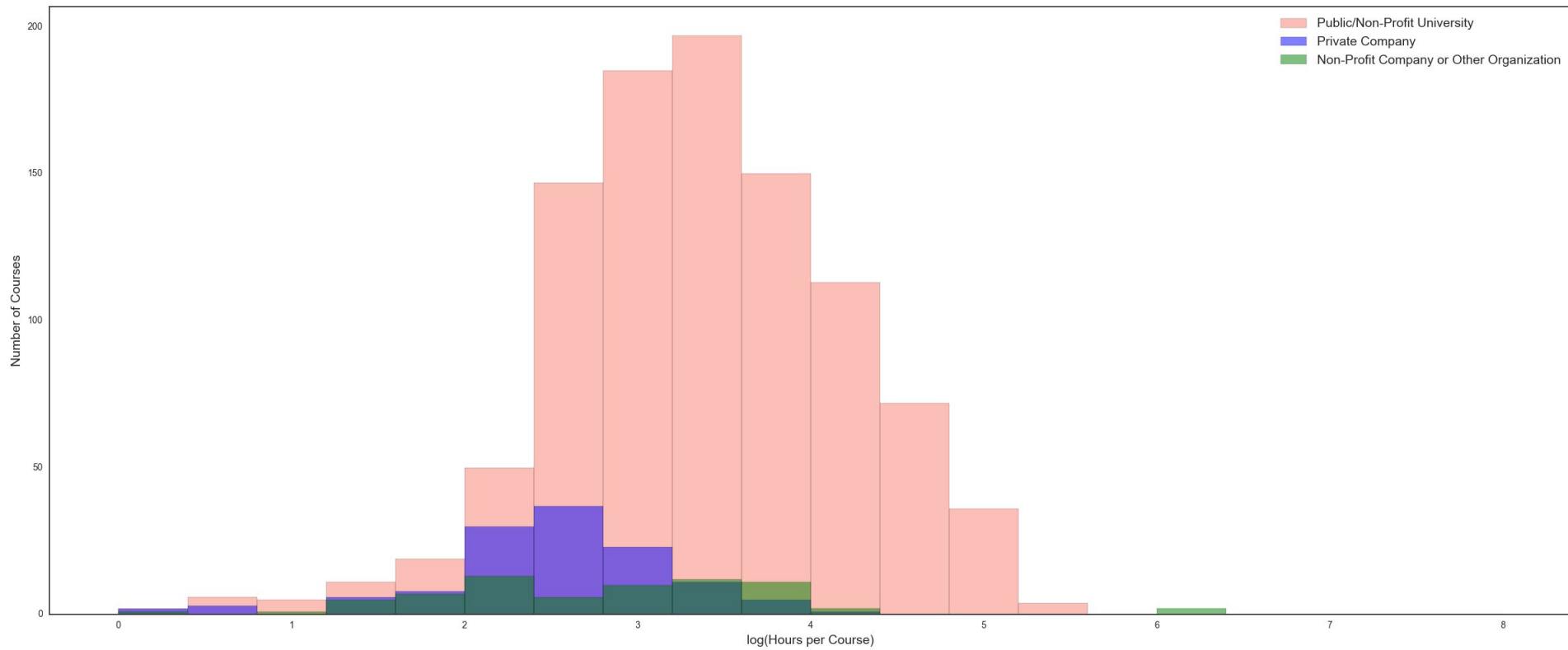
Is there a meaningful difference in course length based on the type of institution providing the course?

Distribution of Course Lengths by Institution Type



Log Transformation - Closer to Normal Distribution

X axis = $\log(\text{expected hours for each course})$



Institution Types - log(Course Length) Summary Statistics

| Type of Institution | Number of Courses Offered | log(Course Length) (mean) | log(Course Length) (var) |
|--|---------------------------|---------------------------|--------------------------|
| Non-Profit Company or Other Organization | 70 | 2.82 | 1.11 |
| Private Company | 126 | 2.53 | .54 |
| Public or Non-Profit University | 997 | 3.38 | .70 |

Log transformation -> closer variances

One-way ANOVA Test

H_0 : edX courses offered by Public/Non-Profit Universities, Private Companies, and Non-Profit Companies or Other Organizations have the same average length.

H_A : one or more of these groups has a different mean course length than the others.

F - statistic = 66.11

P - value = 5.93 e-28

Conclusion: reject H_0 in favor of H_A

Note: group variances of transformed course hour values are not equal, though closer than the non-transformed data.

Next Steps

- Scrape edX over time to get periodic MOOC “market” snapshots.
- When do institutions moving into the market, or into certain subjects?
- How does the typical edX course profile change over time?