The Human Side of Data Science

Southern Stats Chat
Apr 2019

Charlotte Wickham
cwickham@gmail.com
@cvwickham

Slides and links at: https://github.com/cwickham/human-side

There are two sides to our **responsibilities** when we are teaching Data Science

Technical side

We are teaching students to use data to gain knowledge

Computational skills

Statistical thinking

I spend most of my time thinking about this side...

Human side

We are representing who data scientists are and what they value

Our community is diverse and **inclusive**

We care about the **ethical** aspects of our work

and not enough time thinking about this side

Inclusion

short story +
a concrete recommendation

A.K.A. How I ended up in Statistics...

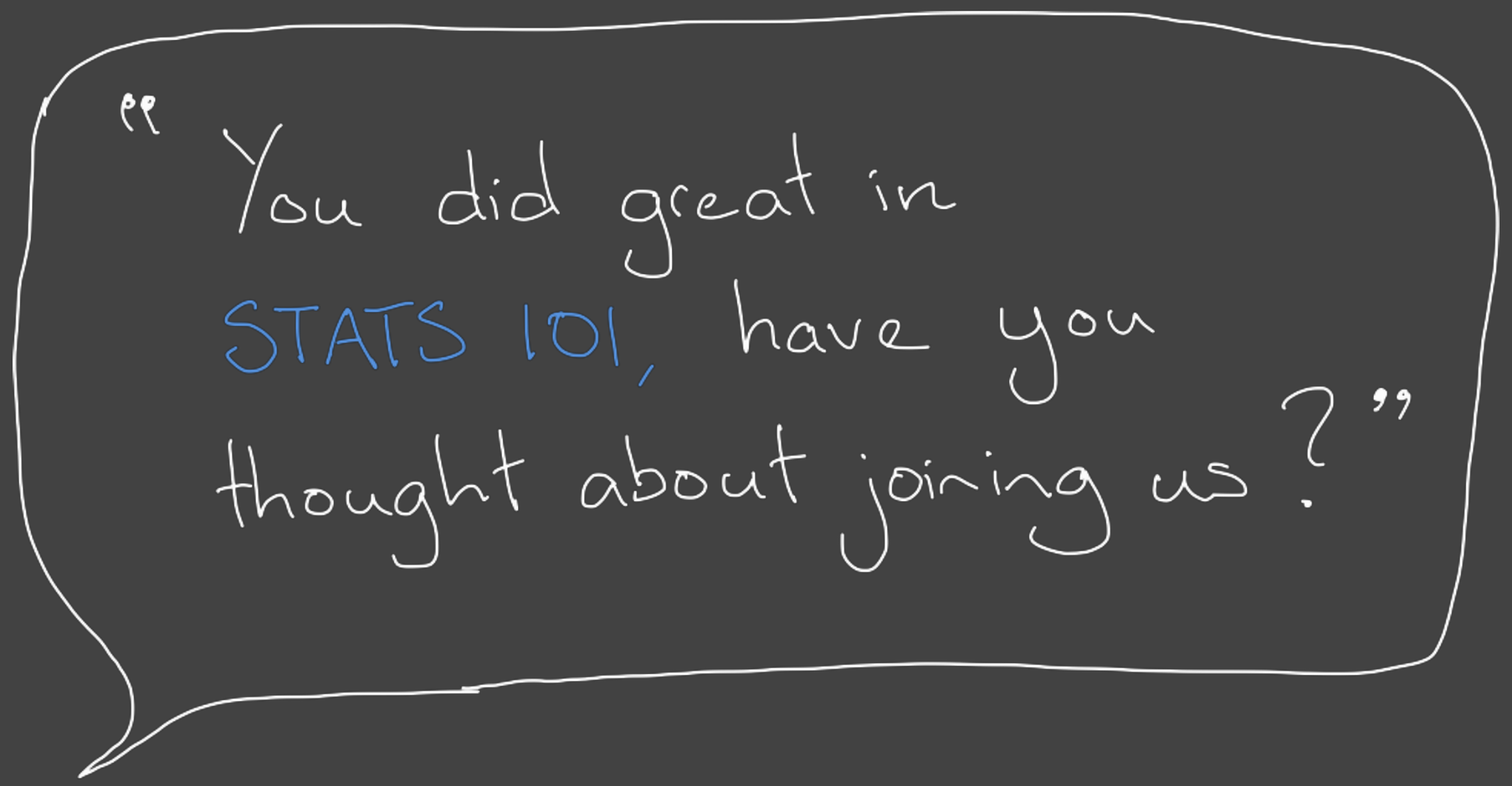
Starting college: Physics and Maths

PHYSICS

MATHS

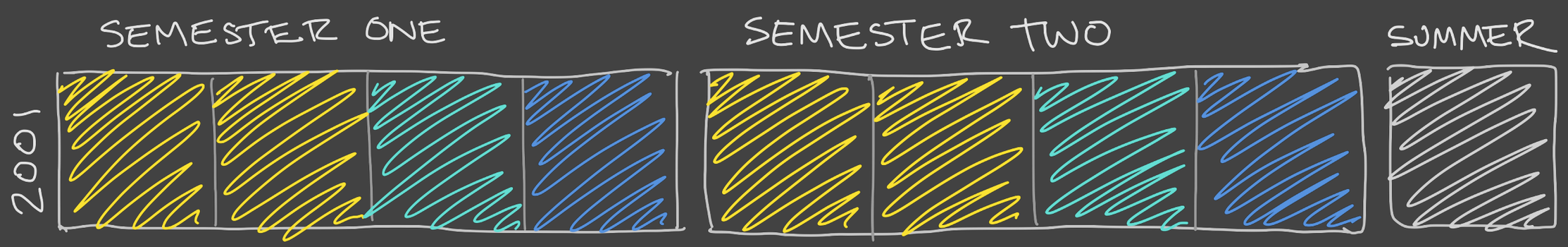
STATS

At the end of my first semester...



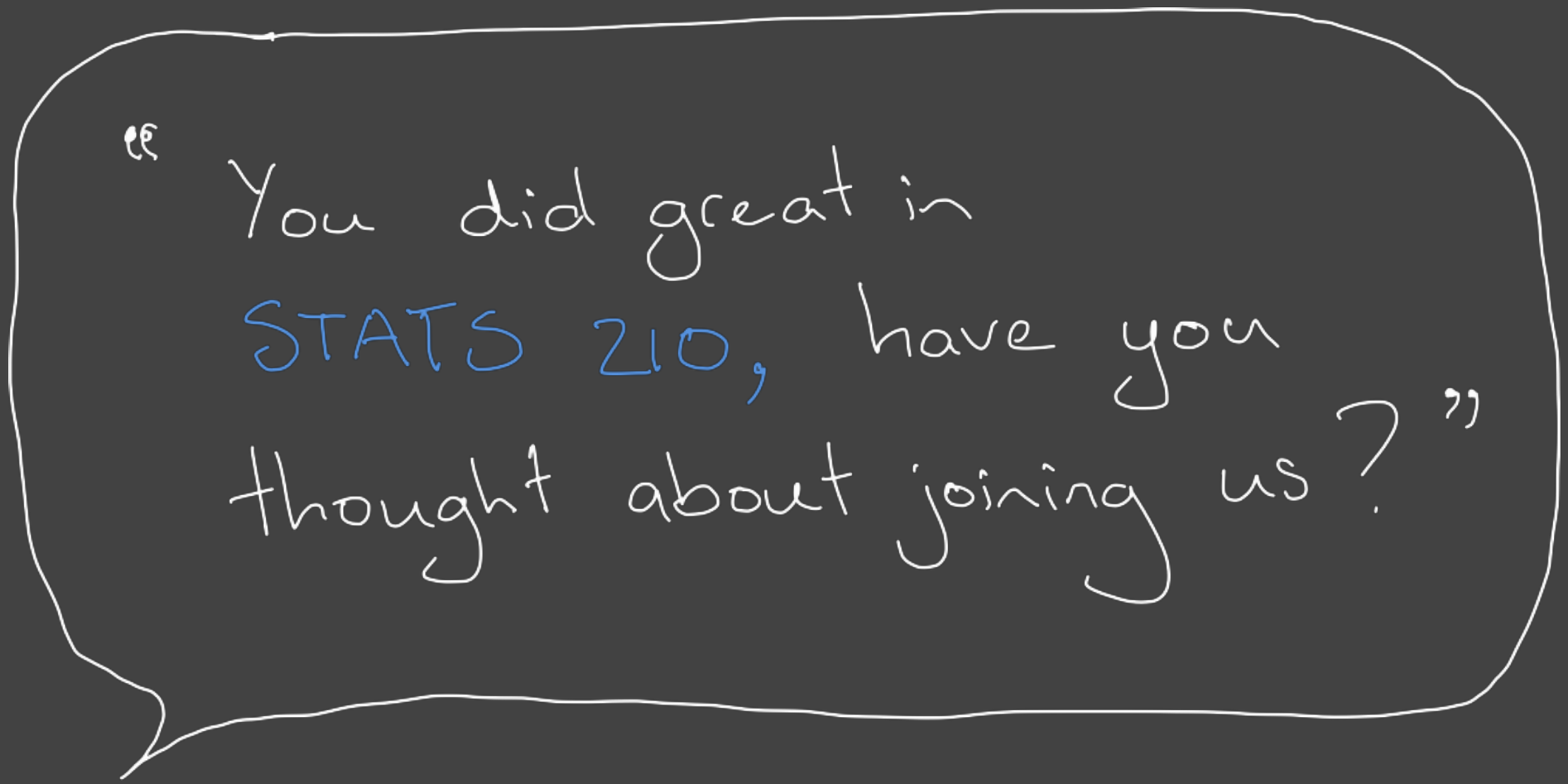
***Paraphrased from memory*

Still thinking: Physics and Maths



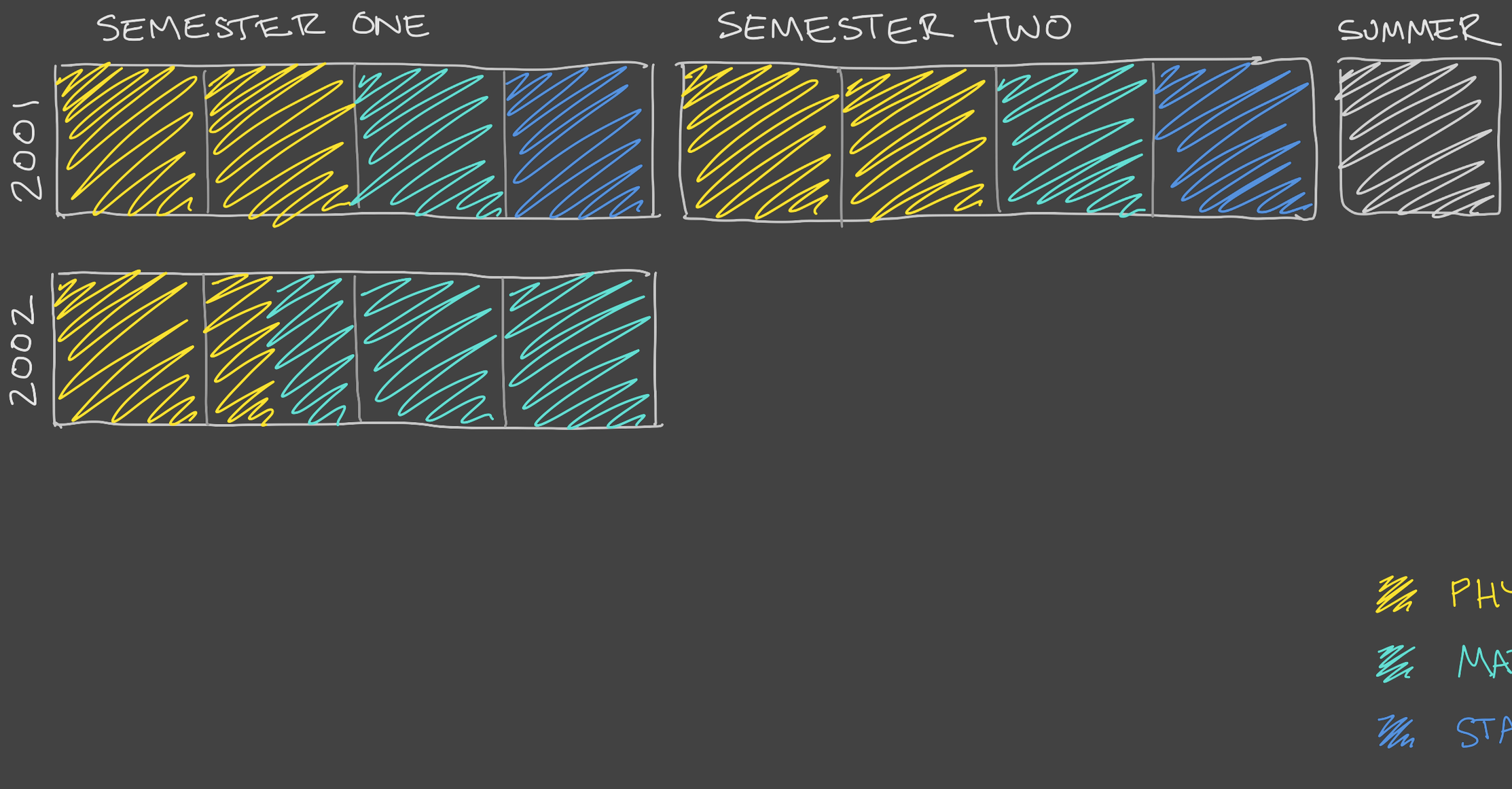
PHYSICS
MATHS
STATS

At the end of my second semester...



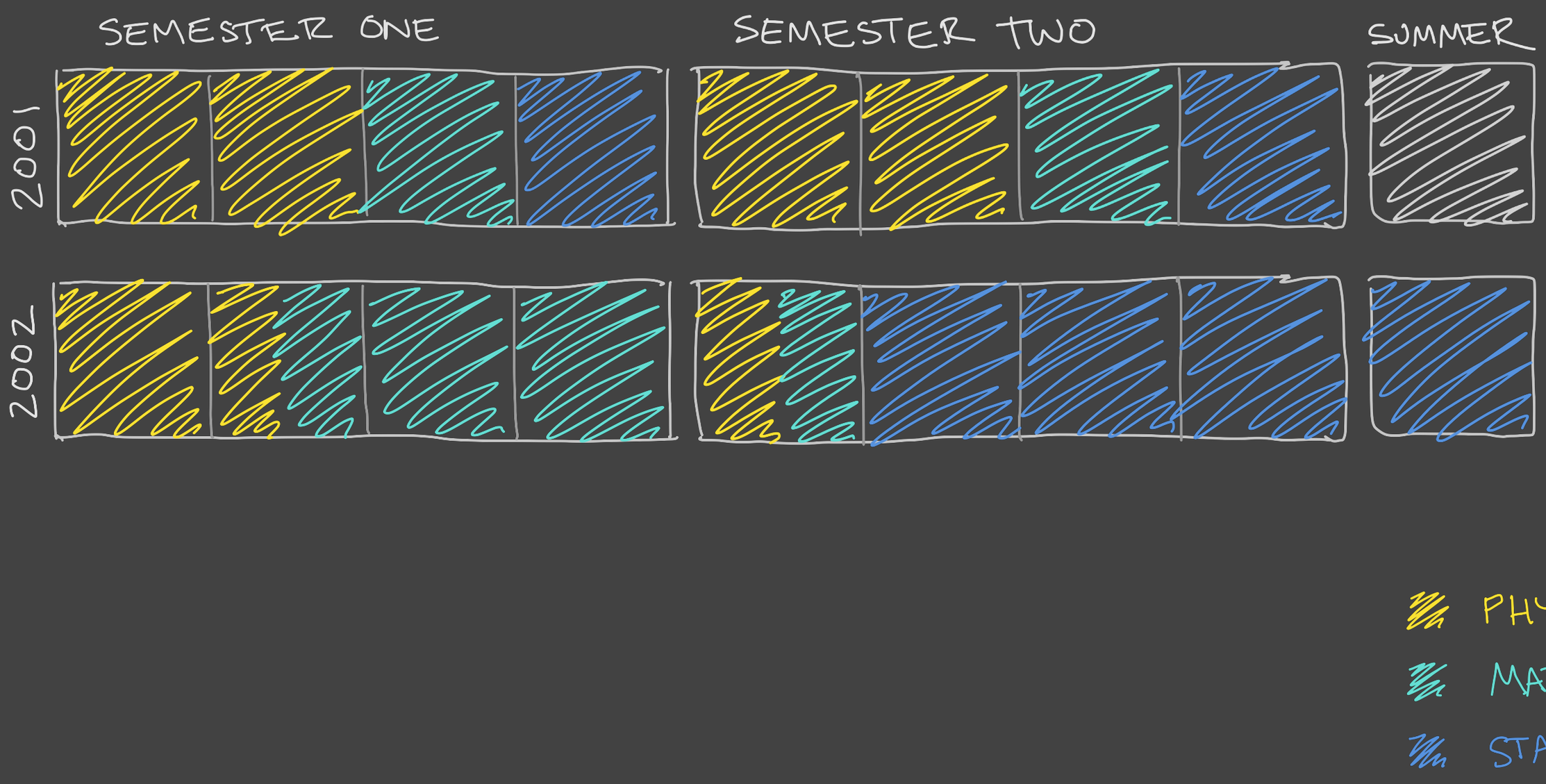
***Paraphrased from memory*

Still thinking: Physics and Maths

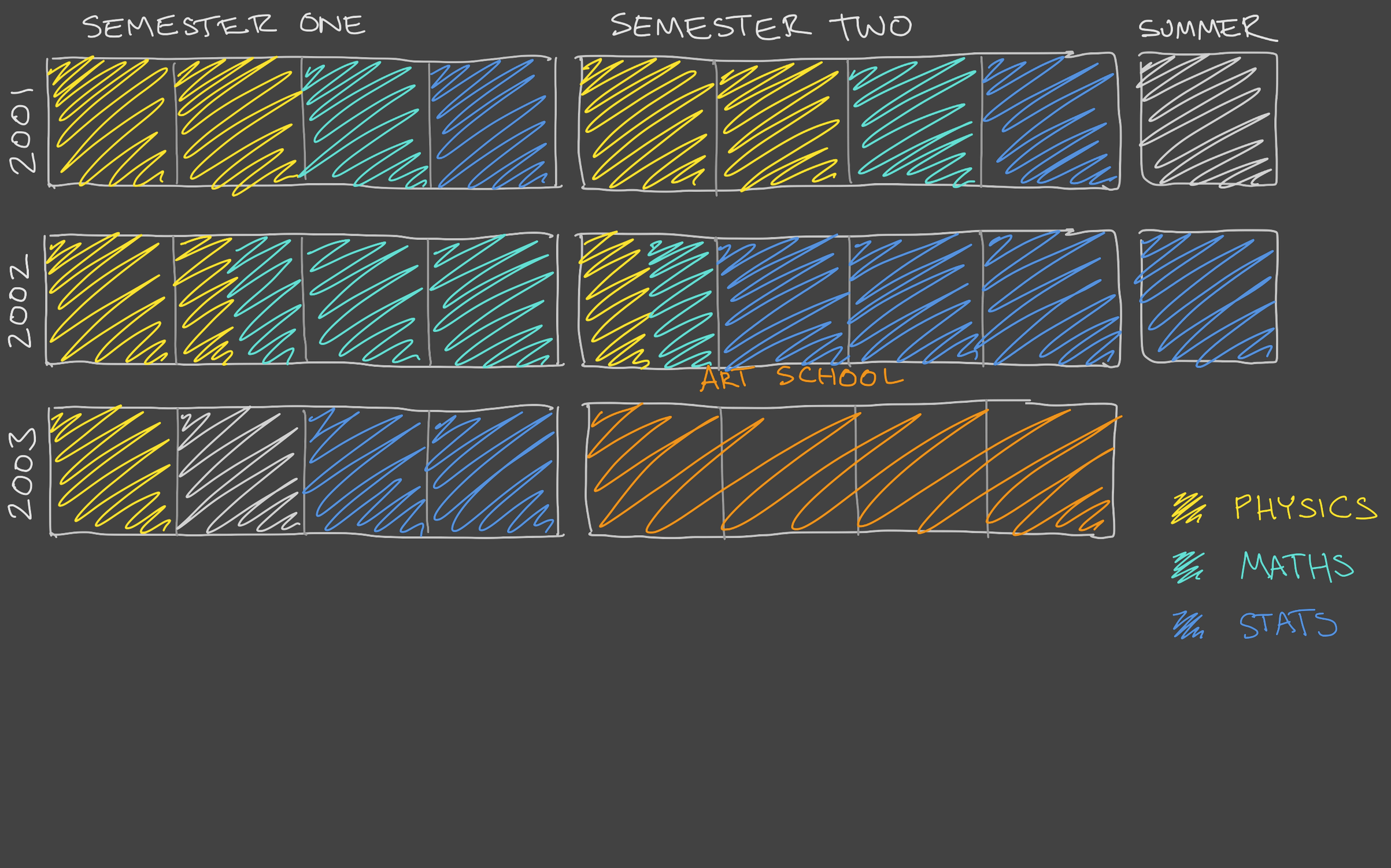


But my thinking was changing...

Now thinking: Statistics



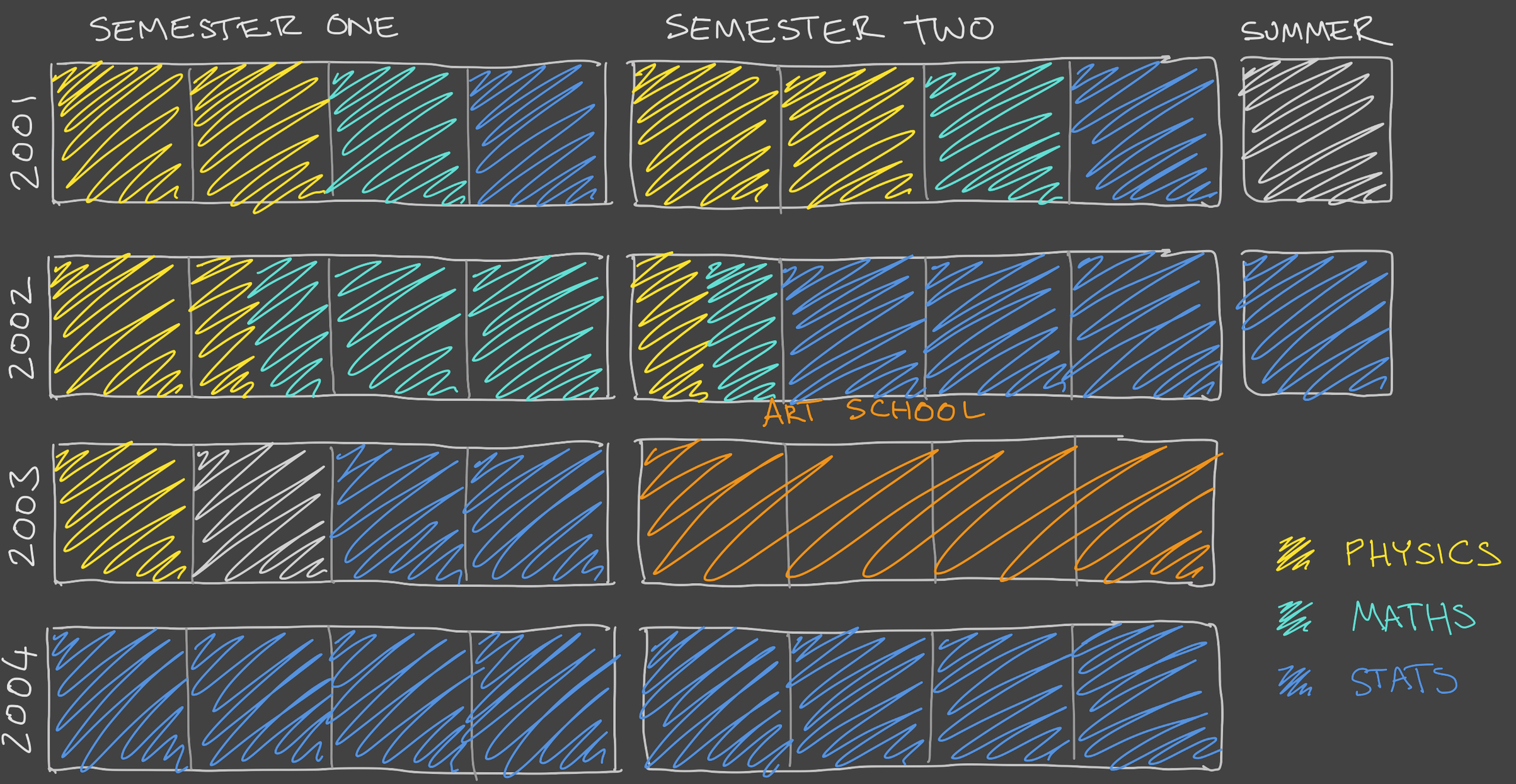
Now thinking: Stats is cool, but what about ...?



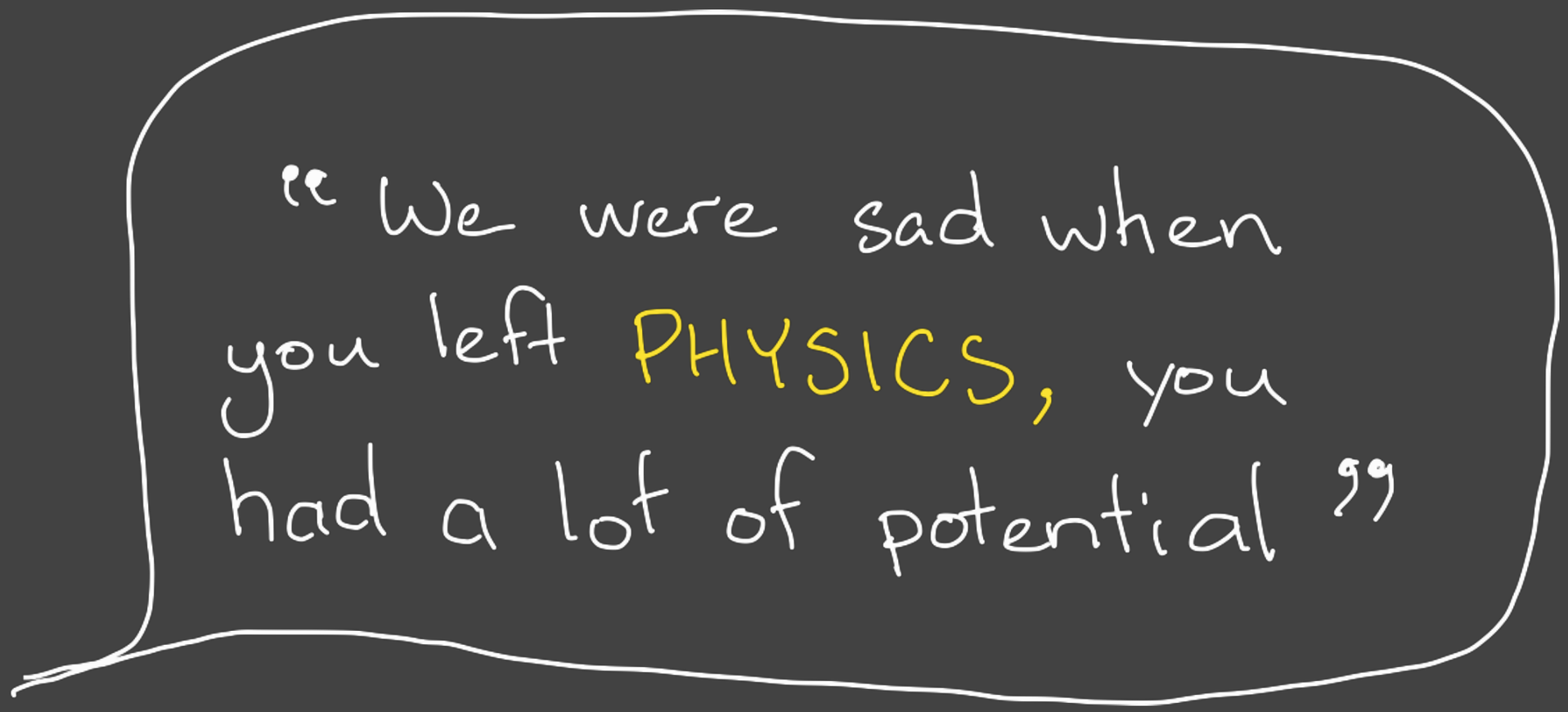
At the end of my third year...

***Paraphrased from memory*

Statistics is for **me**



At graduation I bumped into the Chair of the Physics Department



I had no idea!

***Paraphrased from memory*

What can I do today to create a more inclusive community in CS?

Cynthia Lee

"**Personally** invite a woman or a minority student who did well in your class to major in CS, apply to an internship, or go to grad school. If your TAs work with small groups of students in a discussion section, have them do this as well."

https://docs.google.com/document/d/1hLivou9_wmsZuzKI2pCGQu0KHVIgYfJSaYhvTgO0Wo

What can I do today to create a more inclusive community in ~~cs~~ Data Science?

Cynthia Lee

"**Personally** invite a woman or a minority student who did well in your class to major in Data Science, apply to an internship, or go to grad school. If your TAs work with small groups of students in a discussion section, have them do this as well."

https://docs.google.com/document/d/1hLivou9_wmsZuzKI2pCGQu0KHVIgYfJSaYhvTgO0Wo

What can I do today to create a more inclusive community in ~~cs~~ Data Science?

Cynthia Lee

"This list is designed so you can put it on the wall where you can glance at it from time to time and see **one thing you could work on**. You don't need to do everything at once."

For example

"Take a moment in class today to encourage students to focus on their "slope," not their "y-intercept." That is, in the long run, it matters how fast you're growing and learning, not advantages or deficiencies in where you started."

https://docs.google.com/document/d/1hLivou9_wmsZuzKl2pCGQu0KHVlgYfJSaYhvTgO0Wo

Who are data scientists?

Our students are asking and answering that question based on us and our classes.

We have a responsibility to help them reach the answer:

"Me, I could be a data scientist."

Pick something off Cynthia Lee's list, and do it!

Ethics

short story +
a lot of unanswered questions

How do my pay raises compare to ...?

Salaries at US public universities are public

I'll just get the data
and look...

OSU releases the
"Salary Report"
every quarter...

**as a 1065 page
PDF?!?**

No big deal, I'm a data scientist

	New RStudio project
`library(pdftools)`	Extract text from pdf
`library(stringr)`	Split text into individual records
`library(dplyr)`	Parse individual record text into variables
`library(purrr)`	Repeat for multiple PDFs
`library(dplyr)`	Join by person and job, calculate change in salary
`library(ggplot2)`	Filter, arrange, plot

Make reproducible & transparent

Other people might like to filter, explore and sort this data.

I **should** put all my code and clean data on github.

By making the data easy to access I'm improving transparency.

Something doesn't feel right...

I asked someone I trust...

"Do you think it is ethical to re-publish publicly available, but personally identifiable, data in a form that is easier to search and filter?"

I realized I had an ethical dilemma

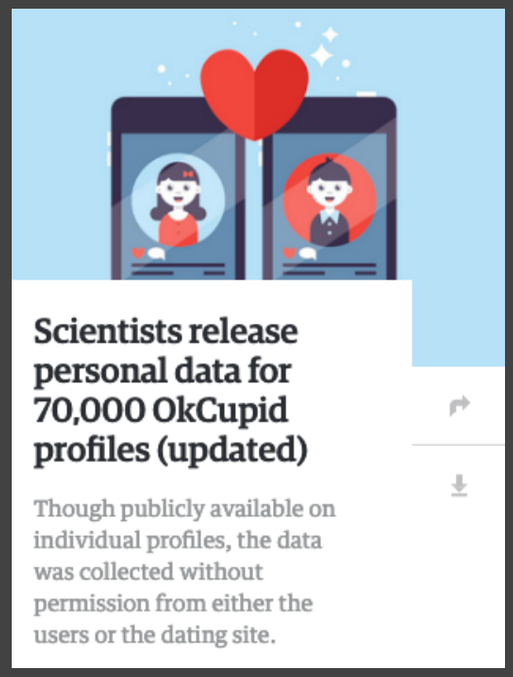
"Do you know of any pointers to good frameworks for helping people make ethical decisions about data?"

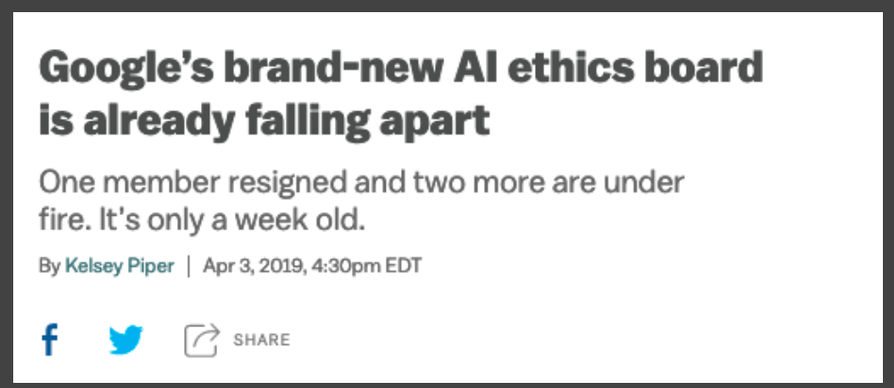
I had no framework *for reasoning about this dilemma*

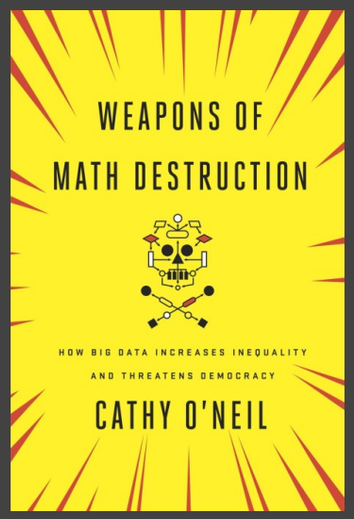
This really bothered me...

It bothered me because,

I'm aware of failures in our field, and believe we need to do better







It bothered me because,

I'm aware of ethical codes in our field,
and endeavor to follow them

- American Statistical Association
  Ethical Guidelines for Statistical practice
  https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx

- The Global Data Ethics Project
  https://www.datafordemocracy.org/project/global-data-ethics-project

- Data Practices - Values and Principles
  https://datapractices.org/manifesto/

It bothered me because,

We value teaching ethical decision making

Goals for Students in Introductory Statistics Classes

Goal 9: "Students should demonstrate an awareness of **ethical** issues associated with sound statistical practice"

Carver, Robert, et al. "Guidelines for assessment and instruction in statistics education (GAISE) college report 2016." (2016).

It bothered me because,

We value teaching ethical decision making

Key Competencies and Features of a Data Science Major

"Programs in data science should feature exposure to and **ethical training** in areas such as citation and data ownership, security and sensitivity of data, consequences and privacy concerns of data analysis, and the professionalism of transparency and reproducibility."

De Veaux, Richard D., et al. "Curriculum guidelines for undergraduate programs in data science." Annual Review of Statistics and Its Application 4 (2017): 15-30.

We value teaching ethical decision making, but

I wasn't taught it

I don't teach this explicitly to my students

I'll share resources I've found most useful

It isn't enough to have memorized a set of ethical guidelines

Guidelines can be in conflict

Guidelines don't cover every situation

Leaves no room for improvement or increasing mastery

Ethical reasoning is **learnable** and **improvable**

Knowledge, Skills and Abilities (KSAs)

Recognize a Moral Issue	Identify decision-making frameworks	Identify and evaluate alternative actions	Make and justify decision	Reflect on decision making

I got stuck here

Rochelle E. Tractenberg & Kevin T. FitzGerald (2012) *A Mastery Rubric for the design and evaluation of an institutional curriculum in the responsible conduct of research*, Assessment & Evaluation in Higher Education, 37:8, 1003-1021, DOI: 10.1080/02602938.2011.596923

There is a learning trajectory

Novice →	Beginner →	Competent →	Proficient
No recognition or only inconsistent recognition of most issues.	Consistent recognition of only most clear-cut issues.	Increased confidence in recognition ability with respect to most, if not all, moral issues.	Identify subtle conflicts at the personal, interpersonal, institutional or societal level. Articulate questions arising either at the level of thought or feeling. Identify moral and ethical components. Analysis of how moral/ ethical question arises. Coherent synthesis of perspectives of all relevant individuals involved for full recognition of moral issues and distinction between moral and ethical issues.

E.g. Recognize a Moral Issue

Excerpt from Table 1

Rochelle E. Tractenberg & Kevin T. FitzGerald (2012) *A Mastery Rubric for the design and evaluation of an institutional curriculum in the responsible conduct of research*, Assessment & Evaluation in Higher Education, 37:8, 1003-1021, DOI: 10.1080/02602938.2011.596923

Professional codes can provide relevant context

ASA Ethical Guidelines Area: **Responsibilities to funders, clients and employers** + public

Recognize a Moral Issue	Identify decision-making frameworks	Identify and evaluate alternative actions	Make and justify decision	Reflect on decision making
Do my actions with respect to data treat one "client" as more important than another? Can I justify prioritizing these responsibilities? Can I rationalize choices made by employers and still maintain professionalism and suitability of my work to the task at hand?	How are my responsibilities to funders, clients and employers with respect to data issues treated under each framework?	Considering my responsibilities to these entities what do my choices about data, its management and sharing necessarily imply?	Are there formal mechanisms by which I can justify my decisions about data its management and sharing? If not, what other justification can I come up with? If so, by what authority does that justification apply to my situation?	What do my choices with respect to data acquisition, management and sharing say about my commitment to funders, clients, and employers? What do they say about my professionalism?

At the **novice** level

Tractenberg, R. E. (2013). *Ethical reasoning for quantitative scientists: A mastery rubric for developmental trajectories, professional identity, and portfolios that document both*. In Proceedings of the 2013 joint statistical meetings, Montreal, Quebec, Canada.

Three dimensions to an ethical curriculum

Knowledge, Skills and Abilities

Level of mastery

Discipline specific context

e.g. ASA Ethical Guidelines

Learning Objectives:
at what level do I want my students to have mastered this KSA?

Learning Activities:
how can I teach this KSA with a relevant example?

Assessment:
what task can I set to assess where my students are in their mastery?

Ethical decision frameworks

the piece that was missing for me

Utilitarian	Which option will produce the most good and do the least harm?
Rights	Which option best respects the rights of all who have a stake?
Justice	Which option treats people equally or proportionately?
Common Good	Which option best serves the community as a whole, not just some members?
Virtue	Which option leads me to act as the sort of person I want to be?

A Framework for Ethical Decision Making
Markkula Center for Applied Ethics at Santa Clara University.
https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/

Ethical behaviour

one piece not addressed by the KSAs

"In a study by Hofling and colleagues, 33 nurses were asked what they would do if a doctor they did not know called them and asked them to give an obviously excessive dose of medicine that was not cleared for use at their hospital."

"31 of 33 said they would not give the medication."

"But when 22 nurses were actually asked to give medicine under these circumstances, 21 were prepared to do so."

Ethical Reasoning ≠ Ethical Behaviour

Prentice, Robert A. *"Teaching behavioral ethics."* Journal of Legal Studies Education, Forthcoming (2014).

What is one thing I can do in my next class?

"The first requirement for **integrating** the ASA Ethical Guidelines into an existing course is to include the Guidelines (see Appendix) in the syllabus."

"...discussion questions/prompts that can be used in any training context."

Tractenberg R.E. (2016) *Institutionalizing Ethical Reasoning: Integrating the ASA's Ethical Guidelines for Professional Practice into Course, Program, and Curriculum.* In: Collmann J., Matei S. (eds) Ethical Reasoning in Big Data. Computational Social Sciences. Springer, Cham

What is one thing I can do in my next class?

Excerpt from Table 1

Text in the ASA Ethical Guidelines	Discussion Questions
Application of these ethical guidelines generally requires good judgment and common sense. In some cases, prioritizing Guideline principles may result in a degree of conflict between different principles; the application of these Guidelines can also depend on issues of law and shared values. **Ethical professional practice in statistics requires following these Guidelines to the extent possible**	List some decisions to which this text might refer. • Identify which principles apply for each of those decisions. • Identify at least two pairs of potentially conflicting principles. • Discuss management and resolution of these two conflicts

Tractenberg R.E. (2016) *Institutionalizing Ethical Reasoning: Integrating the ASA's Ethical Guidelines for Professional Practice into Course, Program, and Curriculum.* In: Collmann J., Matei S. (eds) Ethical Reasoning in Big Data. Computational Social Sciences. Springer, Cham

Questions for discussion

Do you explicitly teach ethical reasoning?

- **Yes**: What do you do?
- **No**: What do you need to help you teach it?