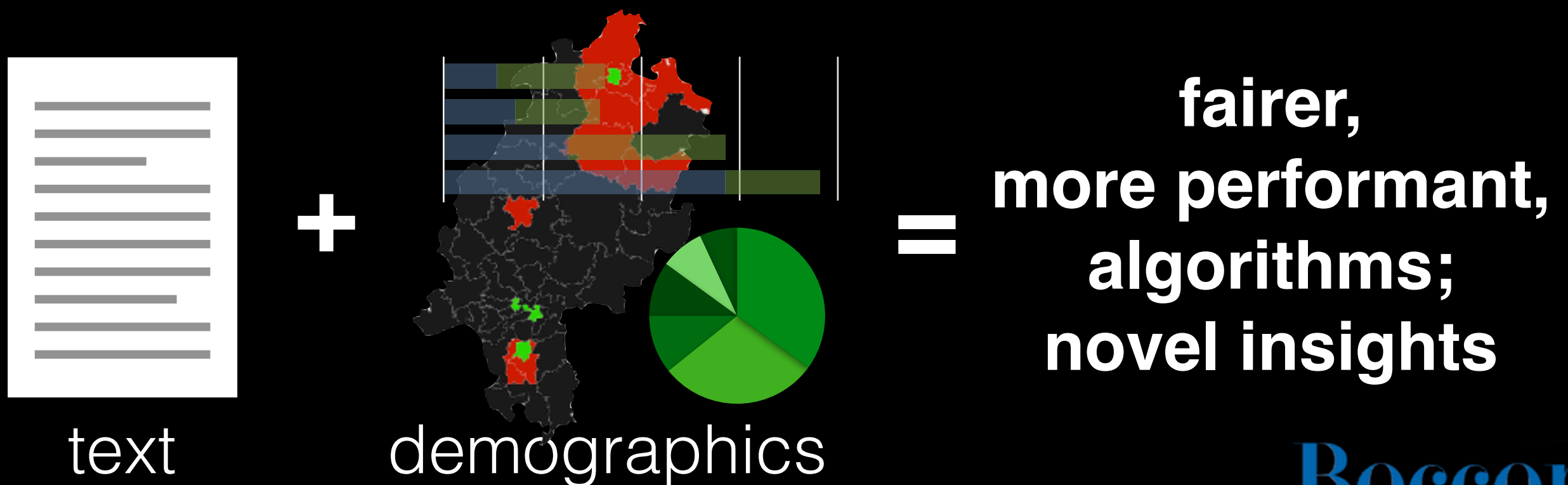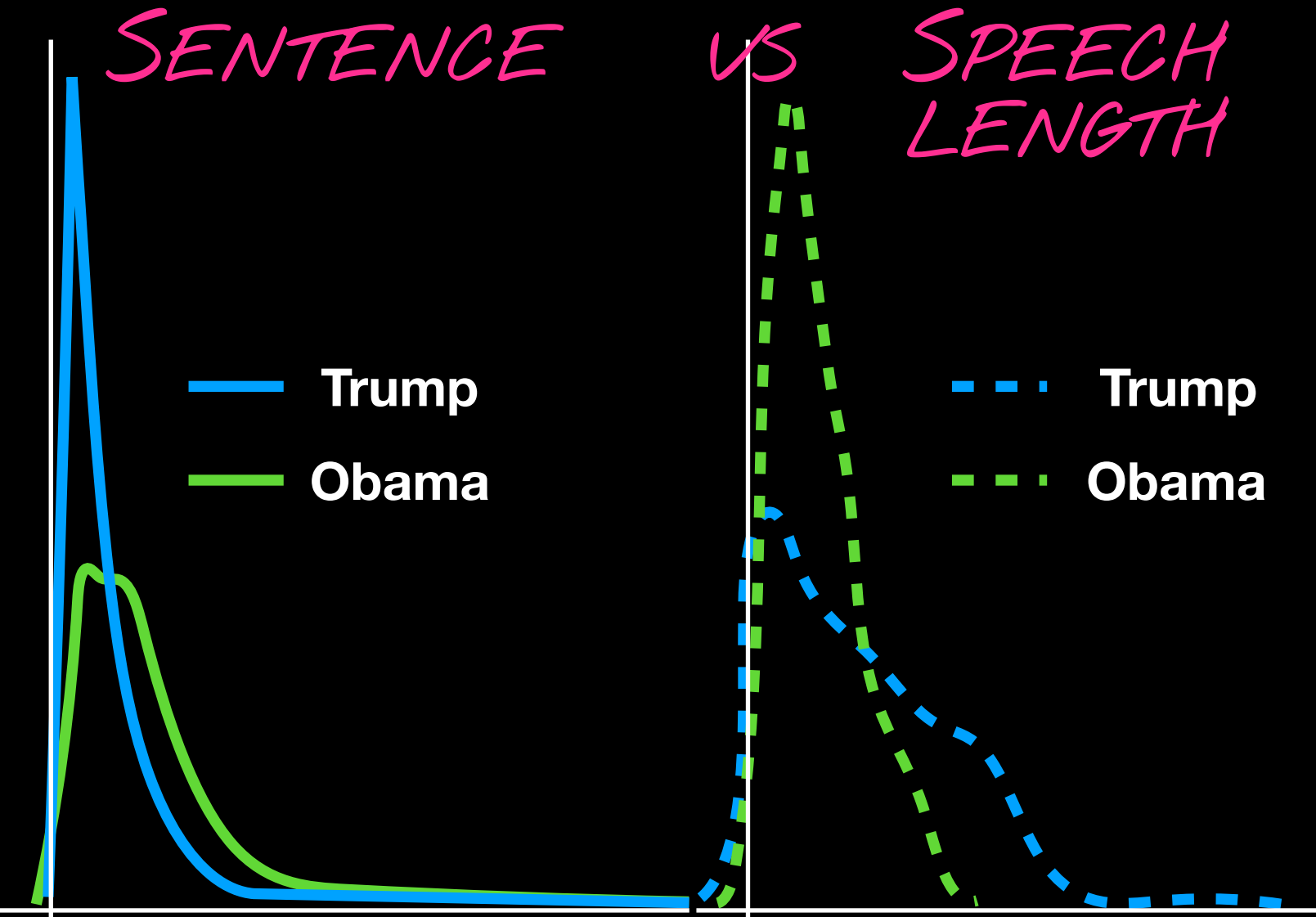# Text Analysis

Spring 2020

Dirk Hovy

New Course, Who dis?

- Master in **linguistics** (Marburg)

- PhD in **Computer Science** (USC)

- Assoc. Prof **Marketing Dept.**; Scientific coordinator **Data and Marketing Insights** research unit (DMI)

Work on NLP, computational social science. **Including demographic information** (Hovy 2015, Benton et al. 2017), **ethics** and fairness (Hovy and Spruit, 2016)

text + demographics = **fairer, more performant, algorithms; novel insights**

Bocconi

Some Projects

SENTENCE vs SPEECH LENGTH
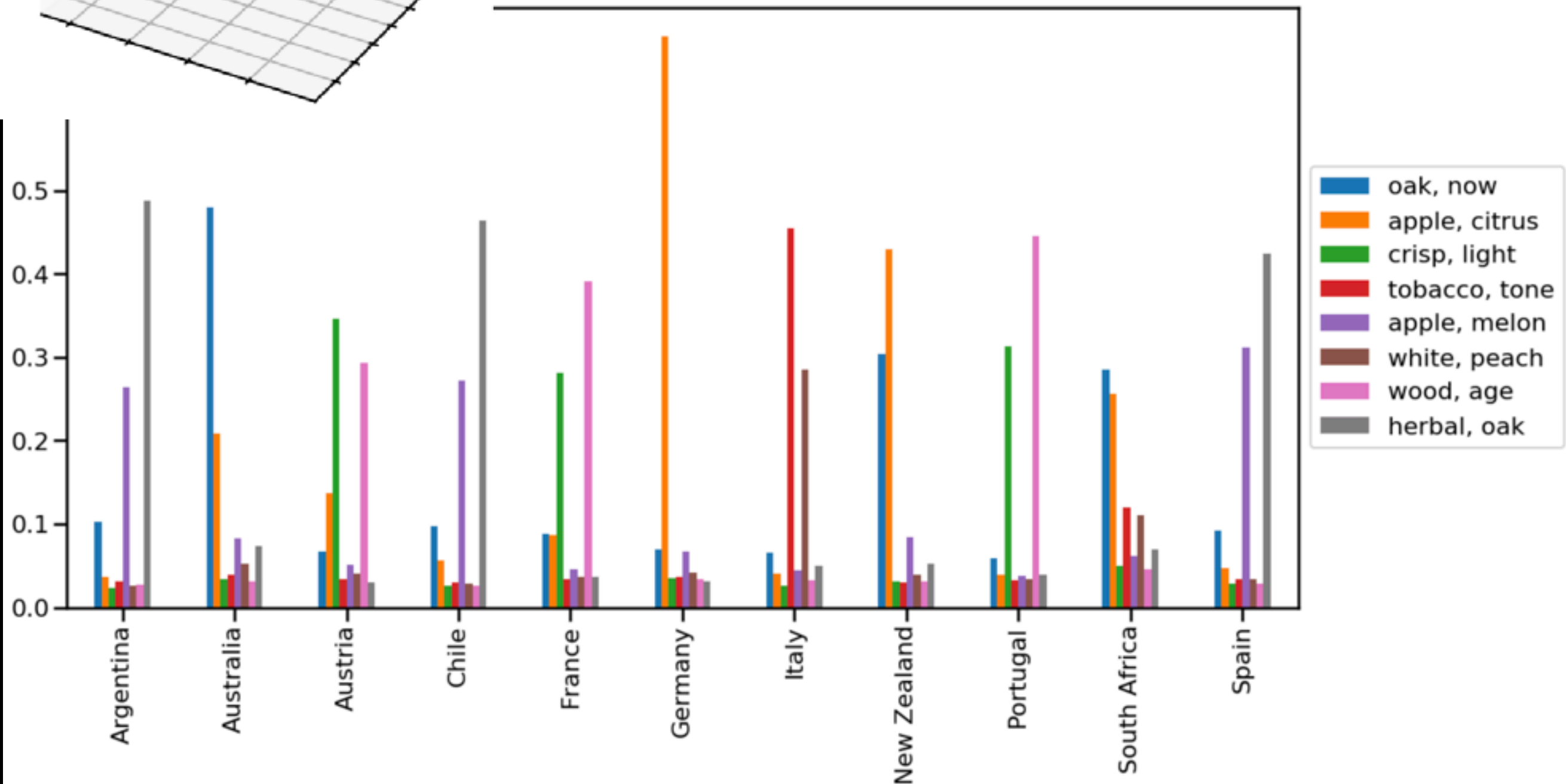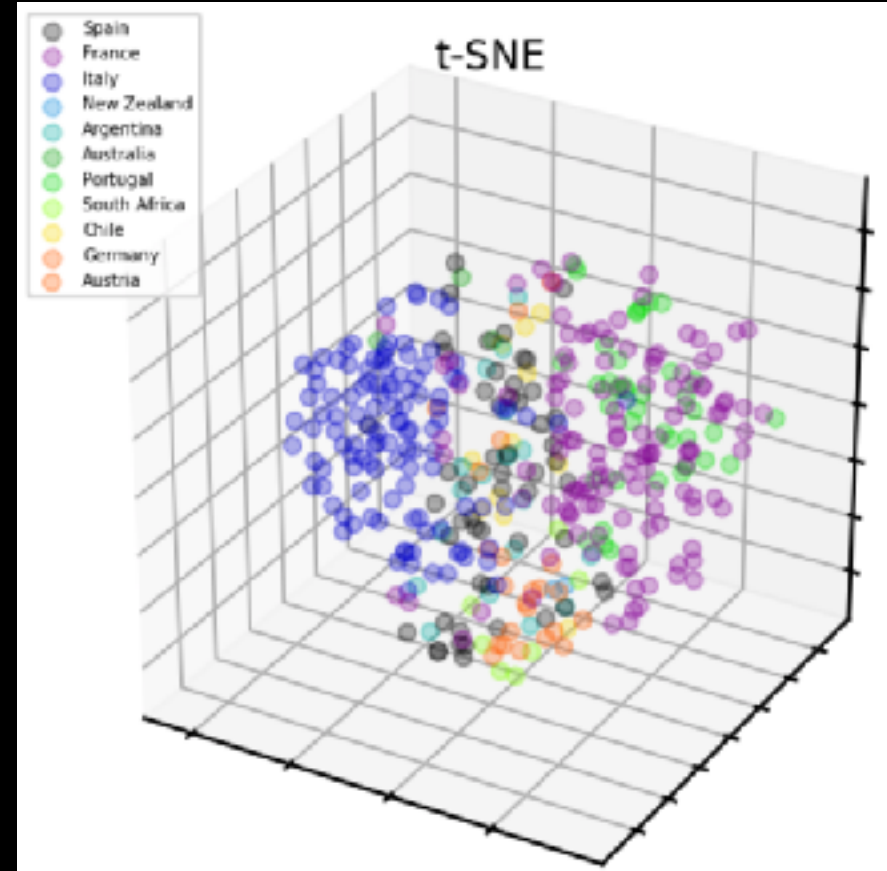
Trump
Obama

Trump
Obama

TALK SIMILARITY

CANDIDATE LMS FINISH A SENTENCE STARTING WITH "AMERICA"

"America first — America first ..."

"America was actually on track to top $ 1 trillion in spending over the coming decade -- because the freedom and dignity --"

3

Bocconi

Some Projects

Wines from different Countries

Course Structure

Part	Content
I. Python Basics	Variables, lists, sets, dictionaries, if-constructs, and for-loops
II. Text Analysis and Machine Learning	Handling and cleaning data, SQL, classifiers

Bocconi

The Bad News

- This might be your hardest class yet

- You won't learn programming by just sitting in class...

Bocconi

The Good News

- You have all the right skills to learn programming

- Practice *does* make perfect

- Successfully running code is immensely satisfying

- Programming literacy is a core qualification for ever more jobs

Bocconi

How do I succeed?

- Follow along: lessons build on each other

- Use the class **notebooks** and bring your **laptop** to class

- **Practice** as we go: do the **exercises**

- **Study** the notebooks: there is more material than just the slides

Bocconi

What If I Get Stuck?

1. **Don't panic**: this happens to everyone

2. Talk to your class mates

3. Google it: `stackoverflow.com` is your friend

4. Come to my office hours (make an appointment ahead of time online)

Bocconi

Grading

Component	Time	Evaluation
Coding Test*	in class, Friday, March 6	50%/100%
Text Classification	online, 1 week	50%/100%

Bocconi

Happy Coding!