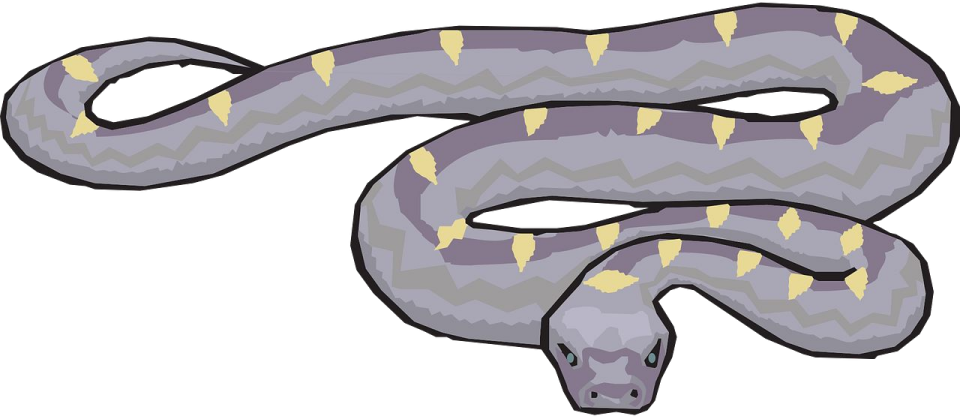
Python for text analysis

Period 1 2019/20



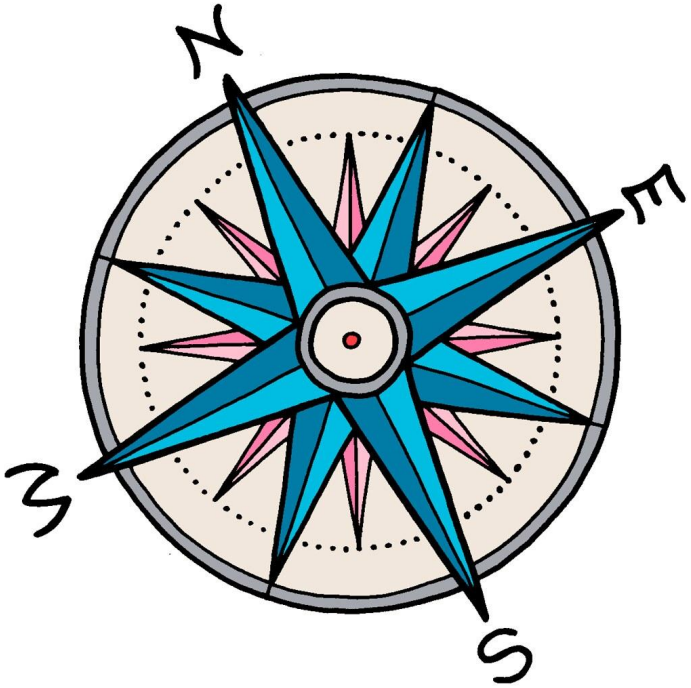
Welcome to python!

python™

Today

- WHO
- WHAT
- HOW
- WHEN

+ Setting up
+ Writing some code :-)



Who?

Teachers

- Marten Postma (m.c.postma@vu.nl)
- Pia Sommerauer (pia.sommerauer@vu.nl)

Teaching assistants

- Anna de Groot
- Sophie Neutel
- Suzana Bašić

Contributors & creators

Also many thanks to these fantastic guys for designing this course and contributing to the material in previous years!

Chantal van Son

Emiel van Miltenburg

Filip Ilievski

Students

- MA Text Mining
- ReMA Human language technology
- Various BA courses

\+ Other interested people (PhD students, research assistants, etc.)

What?

What are you going to lean?

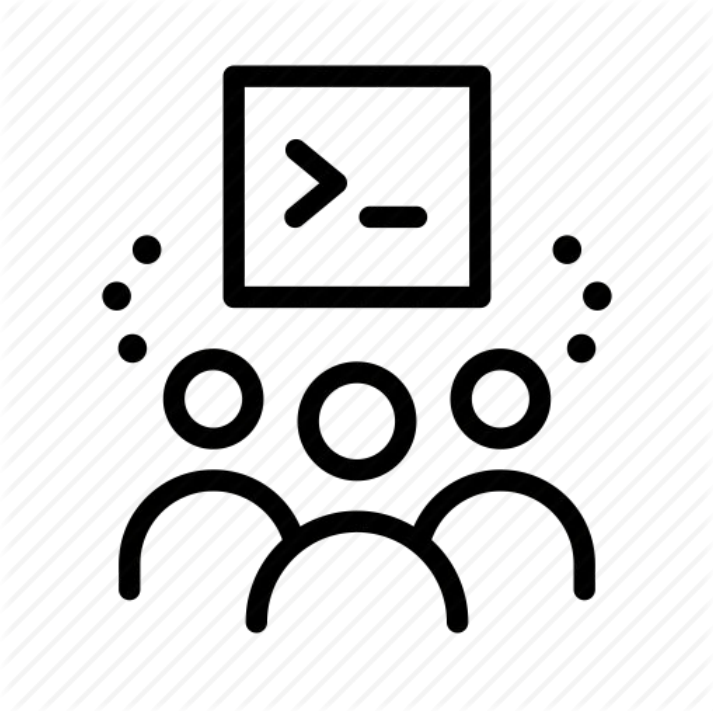
Skills

- Basic python skills (standard library)
- Dealing with some common data structures
- Analyzing text with python
- Document and share code

What are you going to lean?

Core principles

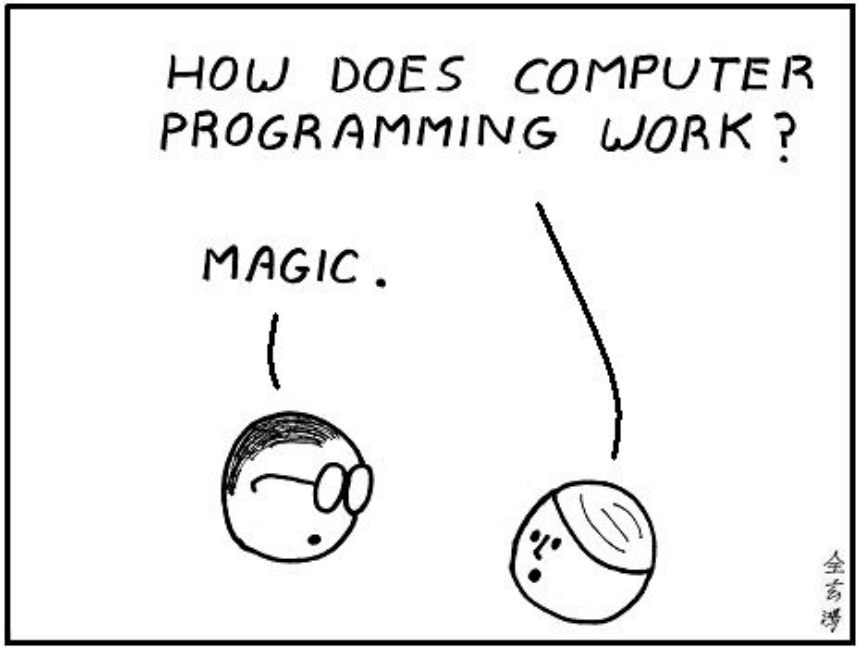
- Readability and transparency
- Problem solving



How?

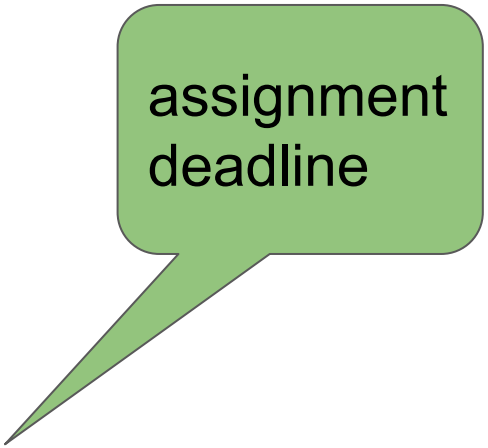
What kind of course is this?

- No prior programming skills required - **anyone can learn programming**
- **Practical course**
- **Participation**



Set-up of the course

Block	
Session 1	new material
Session 2	tutorial
Session 3	feedback on assignment



Set-up of the course

Block 1	Pia
Block 2	Marten
Block 3	Marten
Block 4	Pia
Q&A	Marten & Pia
Exam	

Material

Everything can be found in out Github repository:

https://github.com/cltl/python-for-text-analysis

- **Chapters: introduce material - explanations, examples, exercises**
- **Assignments: Practice & show what you've learnt**
- Additional material: Datasets to play around with, additional exam prep
- README.md: Syllabus - everything you need to know about the course

Grading

Part	**weight %**
Assignments	60
Exam	40
Total	100

Grading - assignments

Part	weight %
Assignment 1	9
Assignment 2	17
Assignment 3	17
Assignment 4	17
Total Assignments	60

A note on ‘getting stuck’

A note on ‘getting stuck’

Strategies for when you get stuck

- Don’t be afraid of errors
- Take a step back from the code - consider the problem
- Class material
- Explain the problem to someone else
- Bring questions to the tutorial sessions
- Take a break

Questions & getting help

If none of the strategies helped, email the teacher in charge of the block.

When

Assignment deadlines

Assignment 1	Friday 09-06-2019 before 23:59
Assignment 2	Tuesday 09-17-2019 before 20:00
Assignment 3	Friday 09-27-2019 before 23:59
Assignment 4	Tuesday 10-08-2019 before 20:00

Exam

Monday 10-21-2019 15.15-18.00

Prep:

- Assignments + in-class work
- Self study (no class): 10-14-2019
- Q&A session: 10-17-2019

Schedule and planning

- Deadlines are strict
- Late submission results in lower grades
- No extensions
- Plan breaks

Detailed schedule: https://github.com/cltl/python-for-text-analysis

Workload: 20h/week

Planning tips

Block		Recommended prep
Session 1	new material	Play around with new material
Session 2	tutorial	Start with assignment and come with questions
	assignment deadline	Finish assignment as well as you can
Session 3	feedback on assignment	Bring remaining questions (if you have them)

Questions?

Setting up

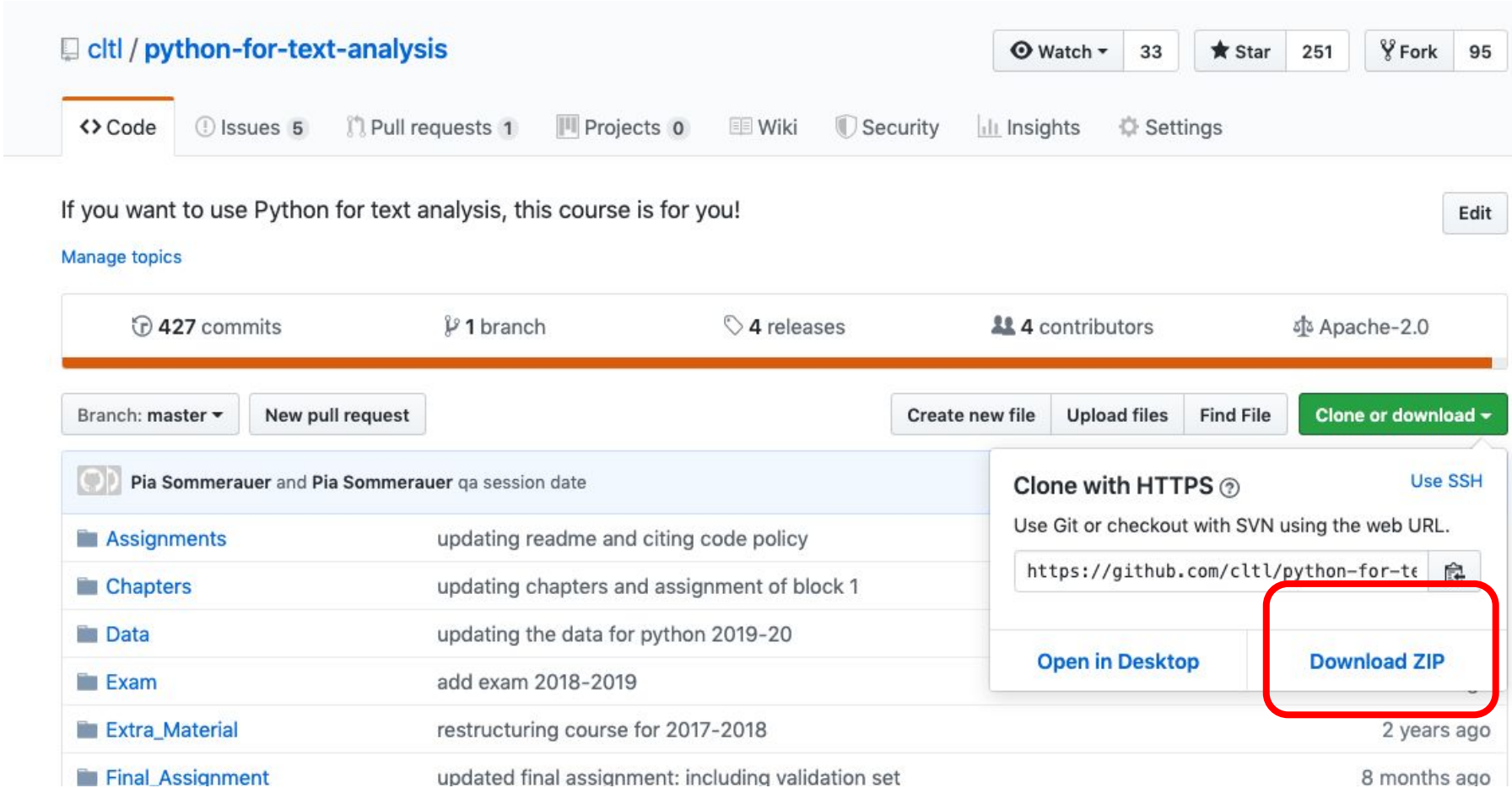
Installing anaconda

https://www.anaconda.com/distribution/

Python 3.7 (most recent version)

Downloading the material

https://github.com/cltl/python-for-text-analysis


cltl / python-for-text-analysis
Watch 33
Star 251
Fork 95
Code
Issues 5
Pull requests 1
Projects 0
Wiki
Security
Insights
Settings
If you want to use Python for text analysis, this course is for you!
Edit
Manage topics
427 commits
1 branch
4 releases
4 contributors
Apache-2.0
Branch: master
New pull request
Create new file
Upload files
Find File
Clone or download
Pia Sommerauer and Pia Sommerauer qa session date
Assignments updating readme and citing code policy
Chapters updating chapters and assignment of block 1
Data updating the data for python 2019-20
Exam add exam 2018-2019
Extra_Material restructuring course for 2017-2018 2 years ago
Final_Assignment updated final assignment: including validation set 8 months ago
Clone with HTTPS
Use SSH
Use Git or checkout with SVN using the web URL.
https://github.com/cltl/python-for-te
Open in Desktop
Download ZIP

Downloading the material

- Store the zip file in a place **where you can find it again**
- Unzip the file

Jupyter notebooks

Windows

> click on the jupyter notebook icon in the start menu

OR

> open a terminal (cmd) and type: `jupyter notebook`

Jupyter notebooks

Mac

> Open the terminal (search for ‘terminal’) and type:

```
jupyter notebook
```

Opening a notebook

- Navigate to the course material (‘python-for-text-analysis’) folder via the jupyter interface
- Click on a notebook

[demo]

Let’s write some code!

What is python?

- Programming language
- Designed to be very readable and intuitive
- Used for dealing with data

Other programming languages:

- C, C++, C#
- Java
- Etc

What is programming?

~ way to tell the computer to perform various tasks (e.g. calculate something, open a file, print text, etc.)

~ using a formalized language (in contrast to ambiguous and messy natural language)

What makes it challenging?

- Learn a new, formalized language
- Having to be very explicit

E.g.: *Guide someone who is blindfolded across a very messy room.*

(taken from https://www.quora.com/What-are-the-best-metaphors-and-analogies-in-learning-Computer-Science)

Why should you learn programming?

- Let your computer perform tasks it is good at (mainly: counting things)
- Deal with data
- - develop your problem solving skills

Also, it’s fun :-)

Outlook

Block 1

- Variables and values (basic building-blocks)
- Integers and floats (numbers and what we can do with them)
- Strings (most important for text)
- Boolean expressions (conditions)

Next session

Block 1 (chapters 1 - 4)

- THU, September 4 3.30-5.15
- Room: OZW-6A01

Preparation:

- Work through and play with material in chapters 1 - 4
- Start assignment
- **Bring questions**

Pia’s guidelines for getting help

- I’ll check emails as soon as possible, but please note that it may take me up to 2 week-days to reply
- The clearer your question the better your chances of a quick answer ;-)
- Please send me notebooks/.py files rather than screen shots
- If you’re feeling very lost and want to discuss things in person, please make an appointment

Questions?

Thank you!
Happy new academic year :-)