

On many ways to
pronounce “Dirichlet” 😊

DAT 562 TEXT MINING

PROF. YULIA NEVSKAYA



de-re-shi-lay



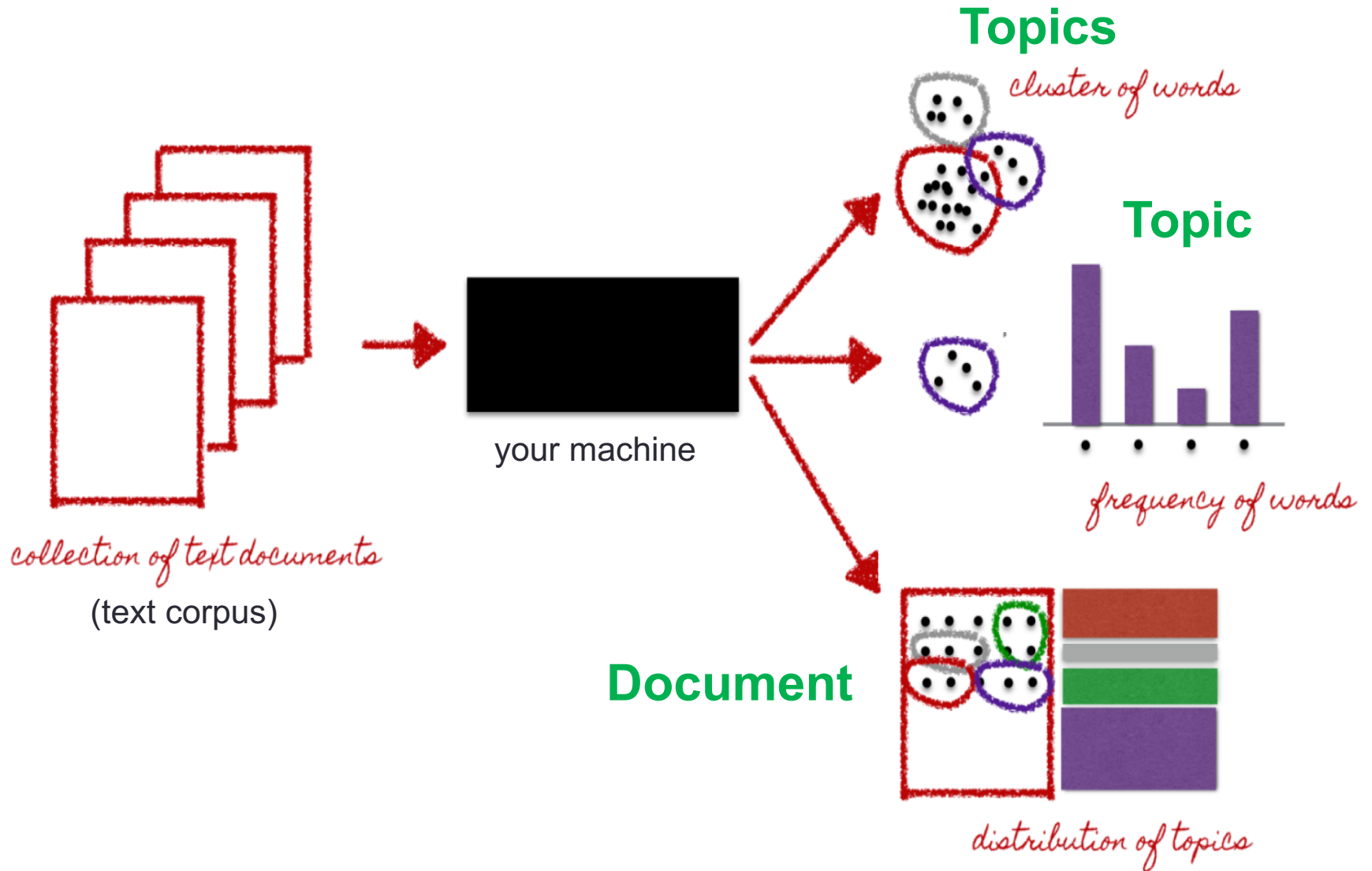
dursh-luh



de-re-clet

- Lab 4: Topic Modeling

Today: Topic Modeling



The Plan for Today

- We'll be doing **topic modeling** in Python

In particular, you will:

0. *Vectorize and normalize the data (of course)*
1. Do a topic model on **toy data** first
2. Model topics on **real data** (data from last lab)
3. Do topic model **visualization**
4. **Evaluate** topic models: *human* judgement versus a *statistic* – who wins?

The Lab Process

- Zoom Breakout Rooms (if you have a remote teammate):
 - Video, audio, chat
 - Share screen
- Your assignment is on Canvas:
 - You **can collaborate** on the assignment with your Zoom lab mates
 - Upload your answer **individually** at the end of the lab (hard deadline is Wednesday by 11.59pm)

What You Need to Do Today

- Go to Canvas / Module 4 / During Lab to get the Assignment
 - Download 2 Jupyter notebooks and put them into the same directory(!) on your machine
- **YOUR TASKS FOR TODAY:**
 - 1) Do all the **EXERCISES** in the script
 - 2) Submit individually the Jupyter notebook with your results and answers on Canvas by the end of the session

Reminders

- Answers for Lab Exercises are available on **Thursdays, 12pm St. Louis time**
- Group Project can't wait! 😊

Jupyter.org/try (Try JupyterLab)

[Install](#) [About Us](#) [Community](#) [Documentation](#) [NBViewer](#)

Try Classic Notebook



A tutorial introducing basic features of Jupyter notebooks and the IPython kernel using the classic Jupyter Notebook interface.

Try JupyterLab



JupyterLab is the new interface for Jupyter notebooks and is ready for general use. Give it a try!

Try Jupyter with Julia



A basic example of using Jupyter with Julia.

Try Jupyter with R



A basic example of using Jupyter with R.

Try Jupyter with C++



A basic example of using Jupyter with C++

Try Jupyter with Scheme



Explore the Calysto Scheme programming language, featuring integration with Python