1. Introduction

Data Management
Spring & Summer 2018
OSIPP, Osaka U

Shuhei Kitamura

About the course

311540/311544 Data Management (M & D)

- Date & time: Tuesdays 5th hour
- Location: Multi-media seminar room (here)
- Instructor: Shuhei Kitamura, Assoc. Prof. in Economics at OSIPP
- Office hours: Fridays 16:00-17:00 @ Toyonaka Sogo 507

About the course (cont.)

Objective: Learn together how to do efficient and replicable empirical research in practice.

In particular, your are expected to:

- Learn methods for doing efficient and replicable empirical research.
  - E.g. How to organize idea, write code, and obtain data.
- Obtain skills to write code in Python and R.

You are also expected to become familiar with useful tools such as Dropbox, Evernote, GitHub, Jupyter Notebook, Readcube, and Slack.

About the class (cont.)

Prerequisite & requirement

- Knowledge in Statistics and Econometrics.
- BYO own laptop
    - OS: Windows
    - Mac/Linux may also work (but I cannot always provide support).

Textbook: No textbook. Relevant references will be provided.

Grading

Two assignments (40%)

- Make and analyze a panel dataset.
- Hand-in your Python/R code (Jupyter Notebook file).

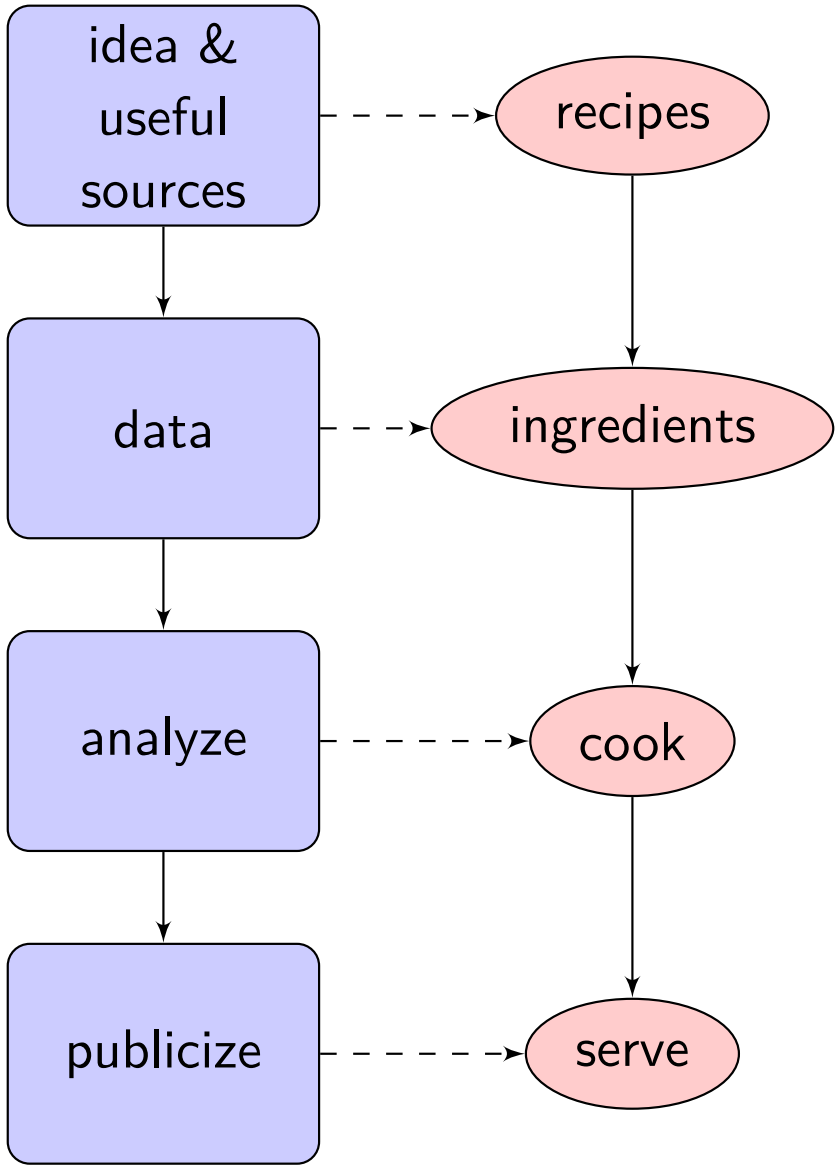
Final paper (50%)

- Write a term paper (max 5 pages).
- You can decide the topic, but you need to obtain data by yourself.
- Hand-in Python/R code, data (if applicable), and the paper.

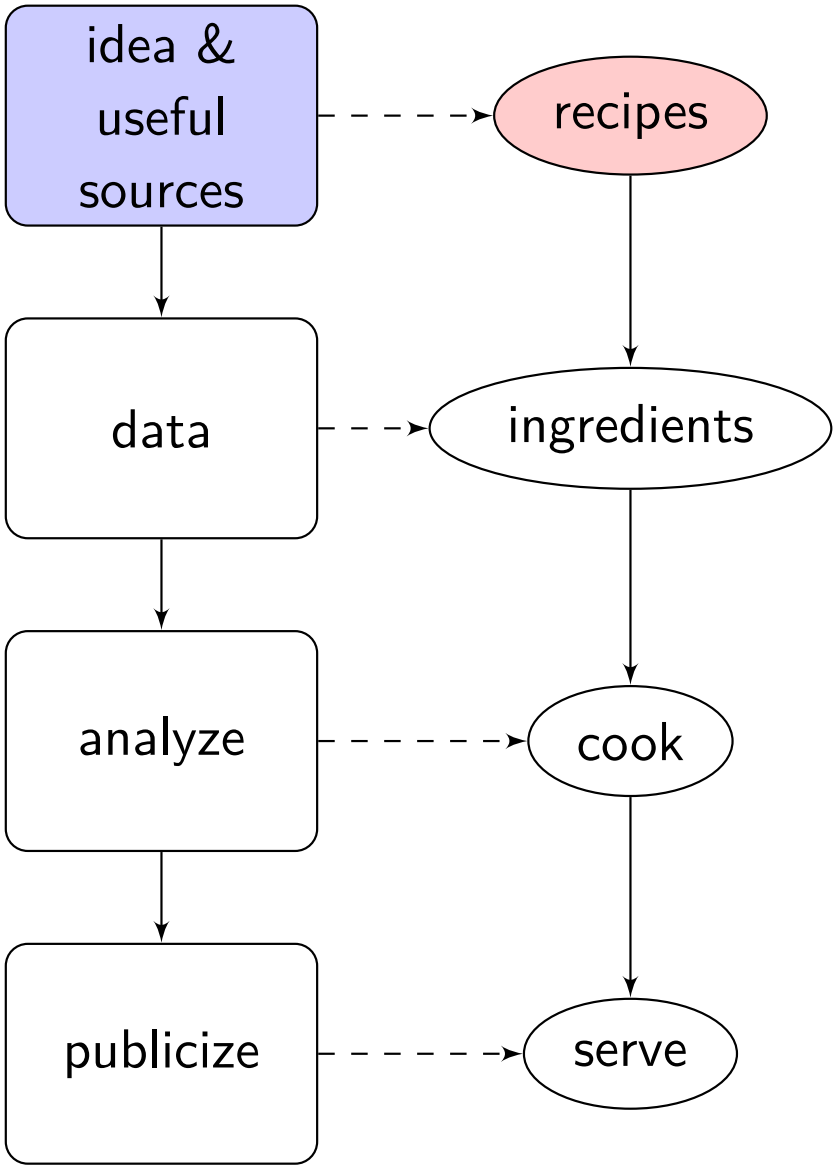
Class attendance (10%)

Notice a slight difference from what has been written in online syllabus.

General workflow of empirical research



Step 1



Good food needs a good recipe.

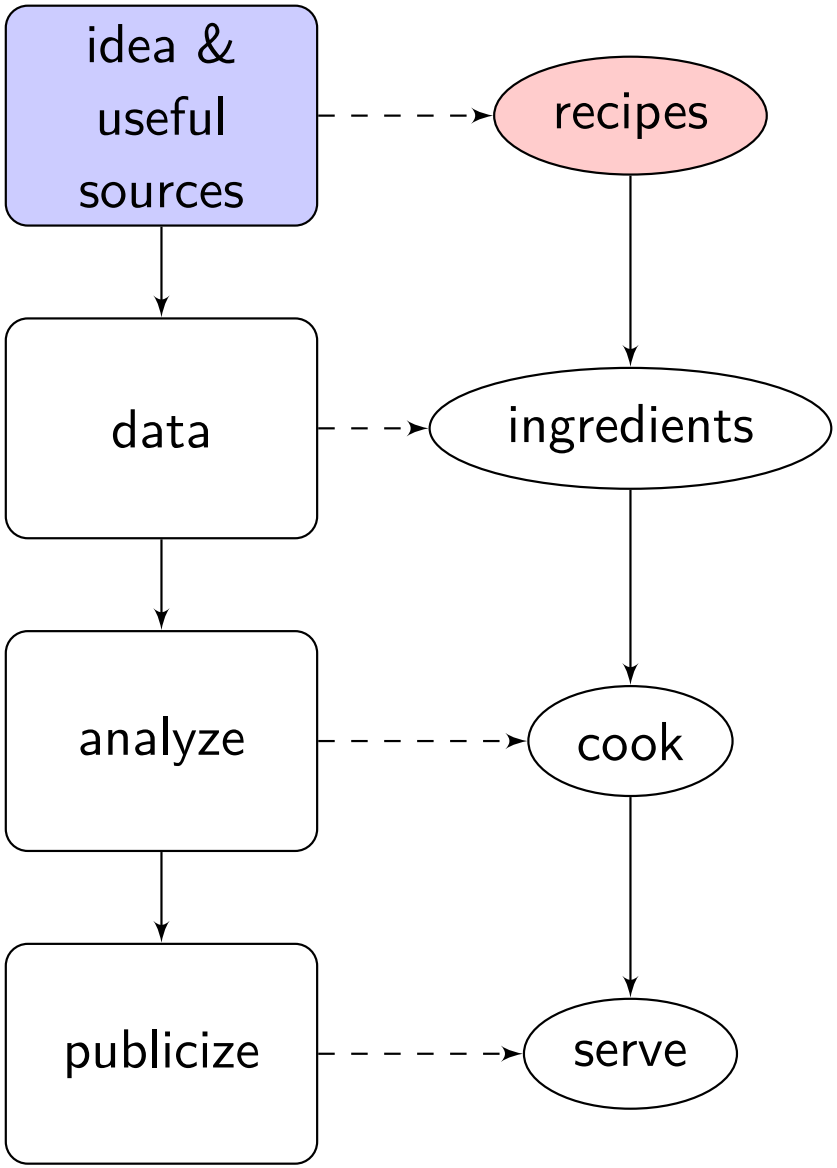
Goal: Get good ideas.

- You may get ideas while reading articles, browsing websites, taking shower, etc.
- You can easily forget the idea or the location where you store it.
- You may combine a new idea with an old one if both are stored in the same place.

Useful tools for storing, organizing, and sharing ideas and resources:

- Evernote
- Readcube
- Dropbox

Step 1



Good food needs a good recipe.

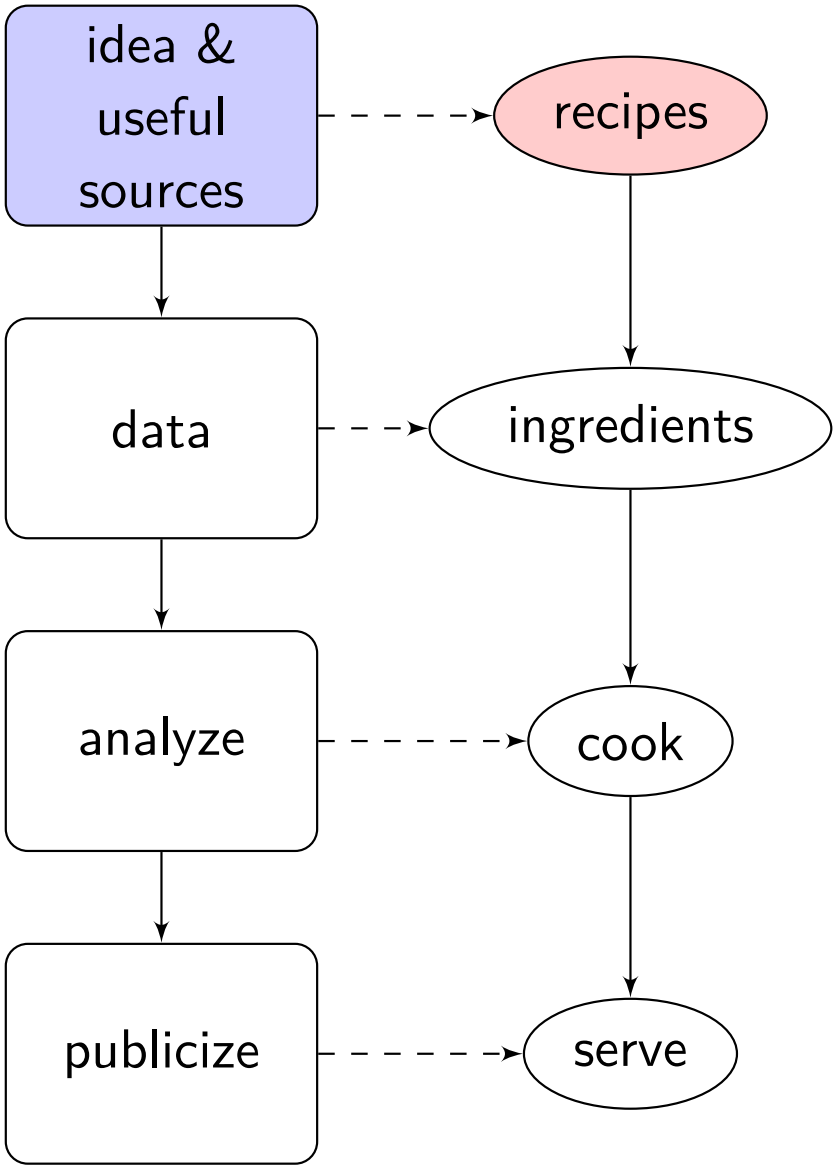
Goal: Get good ideas.

- You may get ideas while reading articles, browsing websites, taking shower, etc.
- You can easily forget the idea or the location where you store it.
- You may combine a new idea with an old one if both are stored in the same place.

Useful tools for storing, organizing, and sharing ideas and resources:

- Evernote
- Readcube
- Dropbox

Step 1



Good food needs a good recipe.

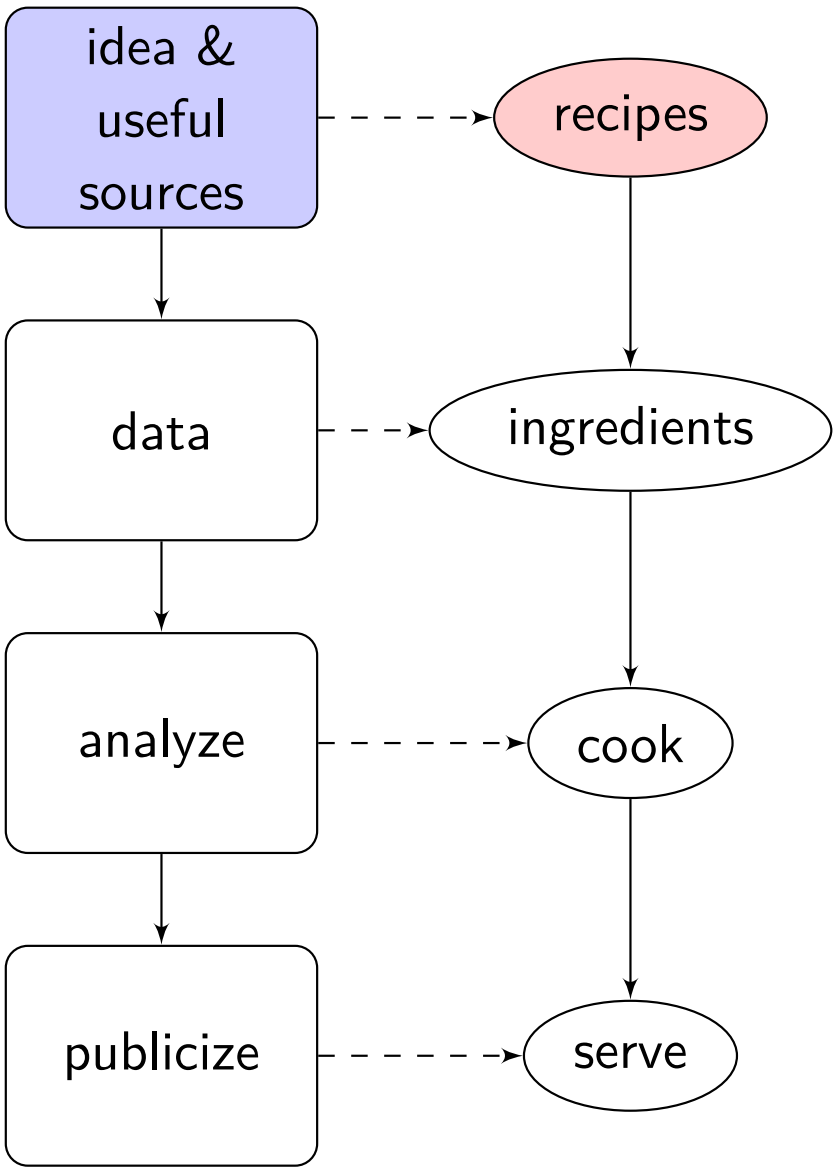
Goal: Get good ideas.

- You may get ideas while reading articles, browsing websites, taking shower, etc.
- You can easily forget the idea or the location where you store it.
- You may combine a new idea with an old one if both are stored in the same place.

Useful tools for storing, organizing, and sharing ideas and resources:

- Evernote
- Readcube
- Dropbox

Step 1



Good food needs a good recipe.

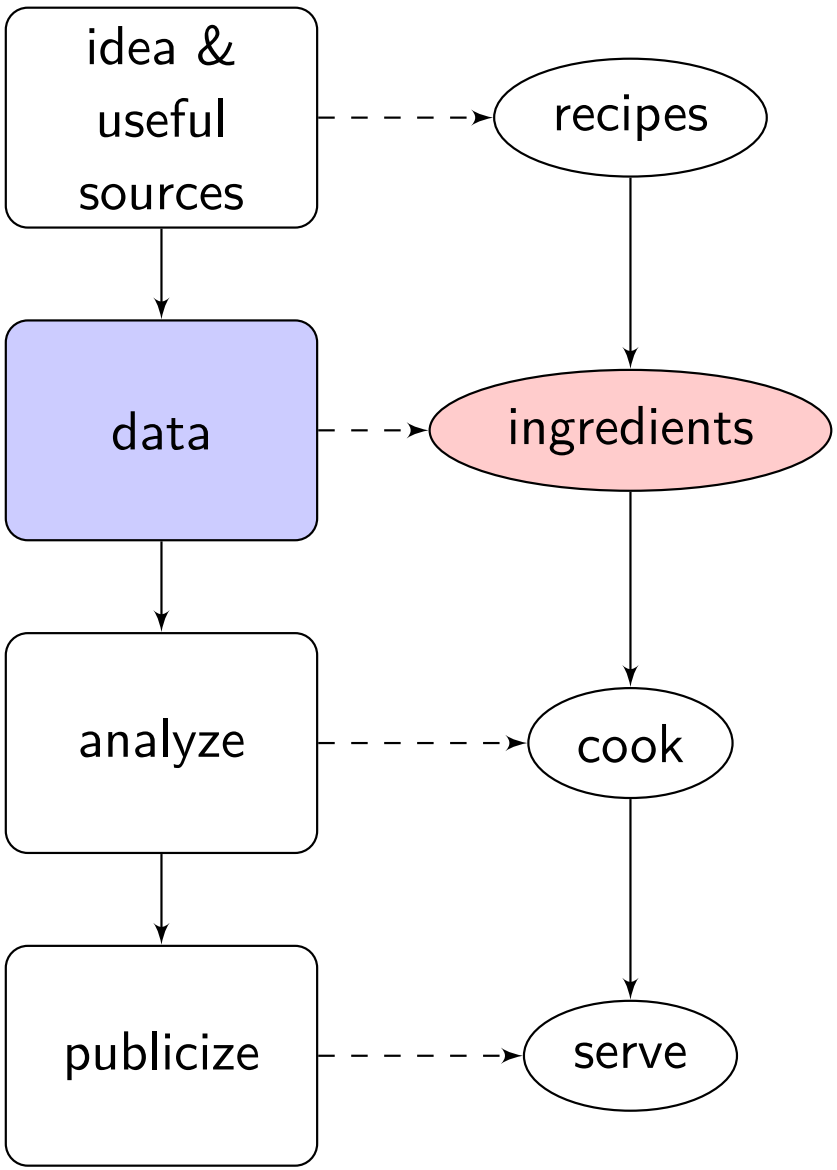
Goal: Get good ideas.

- You may get ideas while reading articles, browsing websites, taking shower, etc.
- You can easily forget the idea or the location where you store it.
- You may combine a new idea with an old one if both are stored in the same place.

Useful tools for storing, organizing, and sharing ideas and resources:

- Evernote
- Readcube
- Dropbox

Step 2



Good food begins with good ingredients.

Goals: Collect good data. Clean them properly.

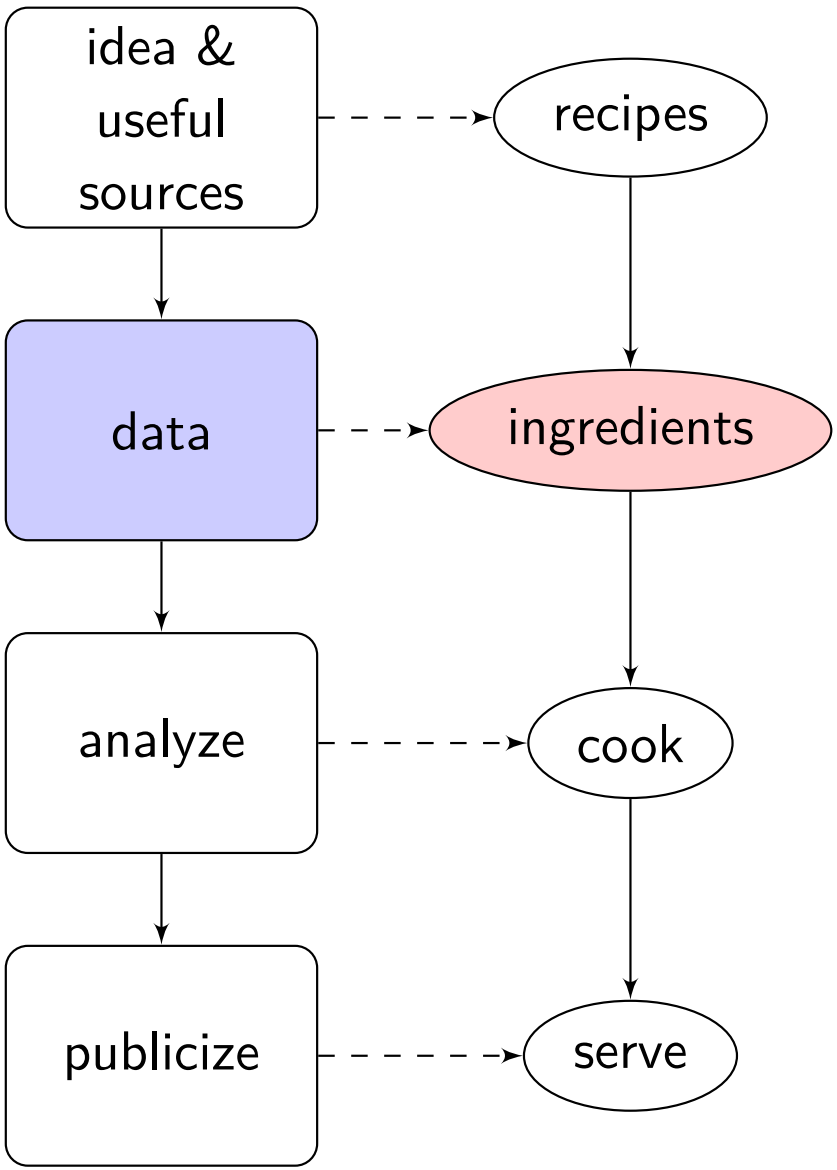
Useful tools for collecting and cleaning data:

- Python & R

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Step 2



Good food begins with good ingredients.

Goals: Collect good data. Clean them properly.

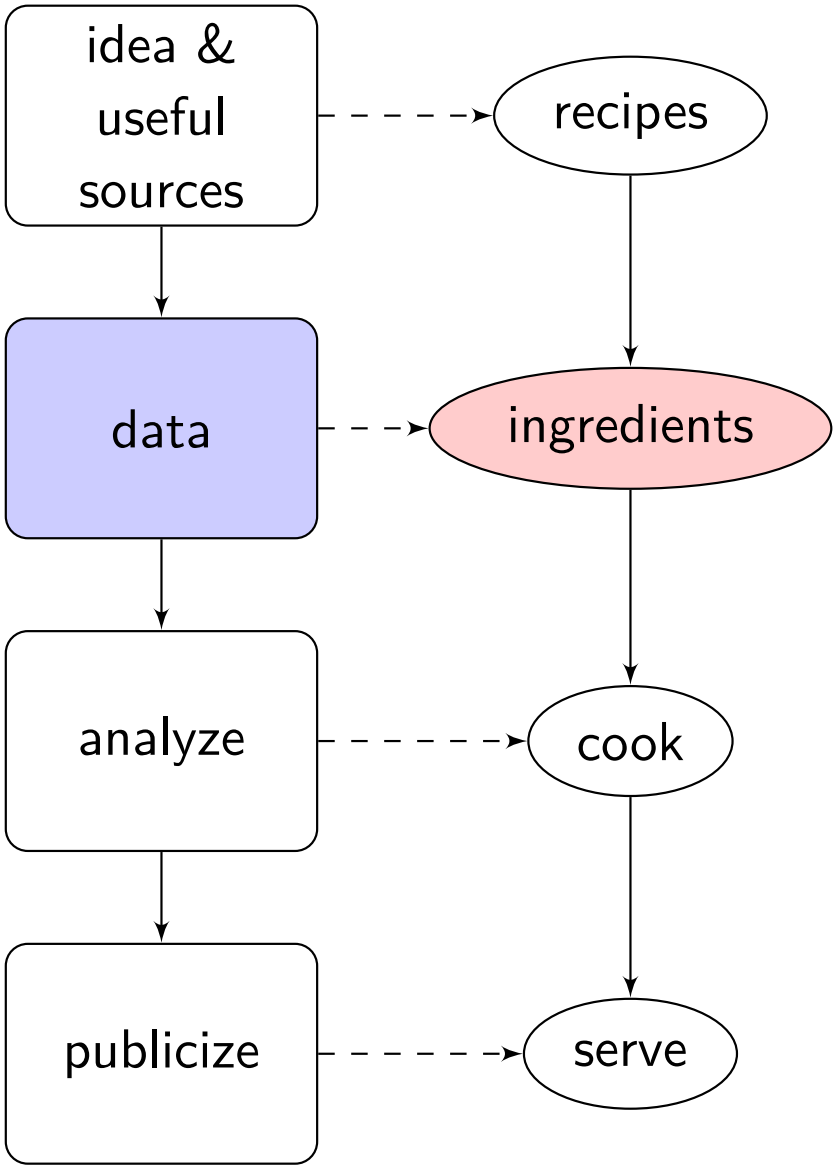
Useful tools for collecting and cleaning data:

- Python & R

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Step 2



Good food begins with good ingredients.

Goals: Collect good data. Clean them properly.

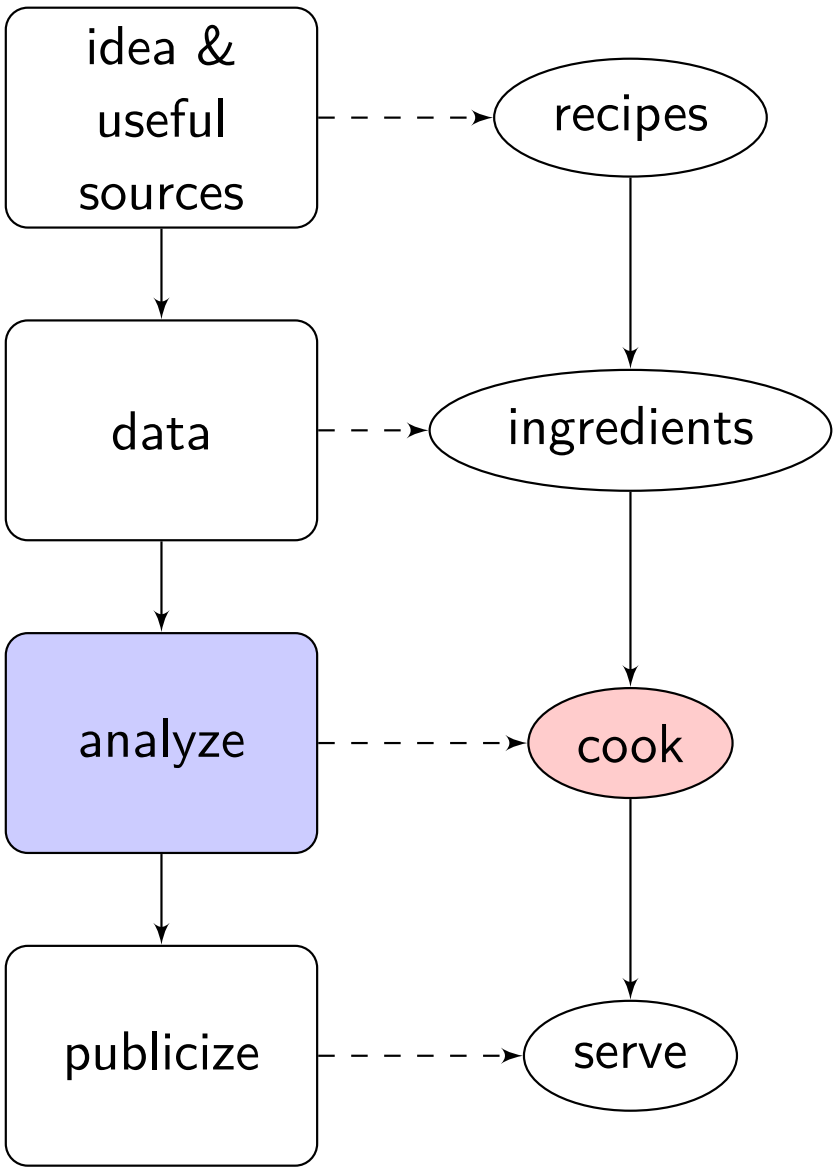
Useful tools for collecting and cleaning data:

- Python & R

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Step 3



Good food is cooked well.

Goal: Analyze data well.

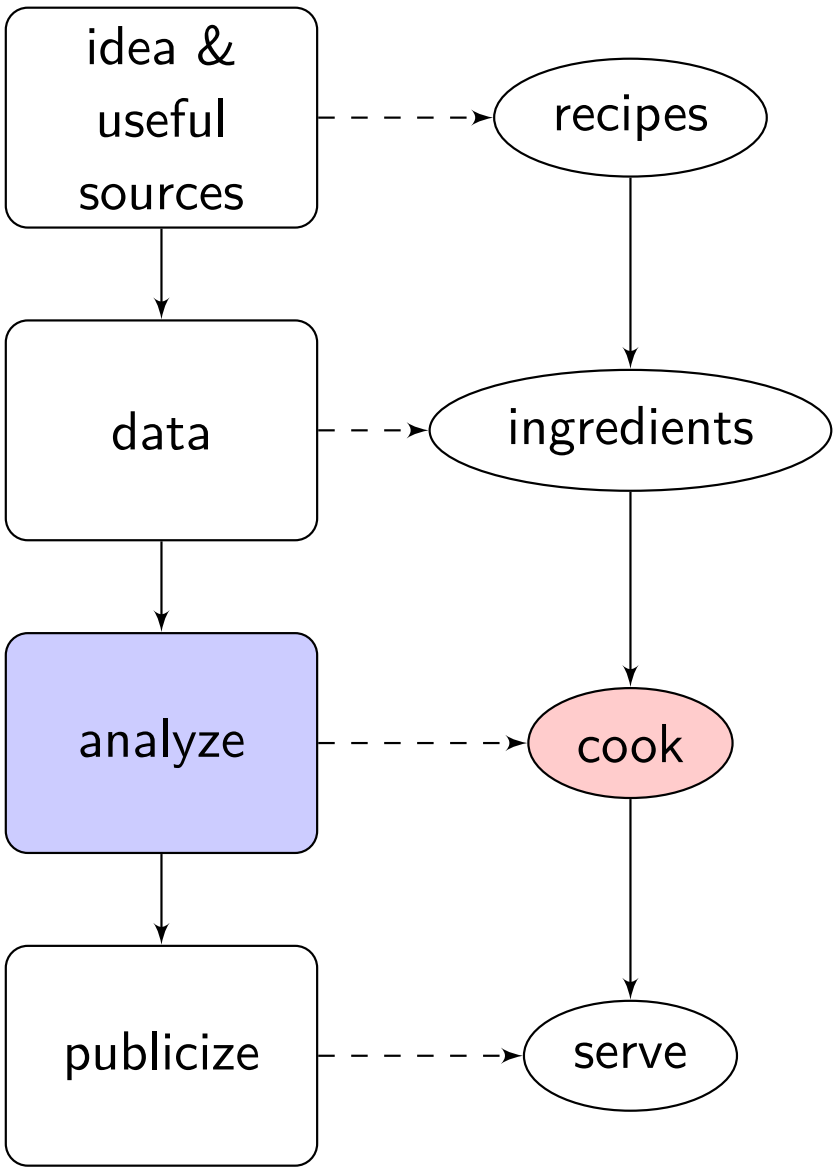
Useful tools for analyzing data:

- Python & R
- (Stata)

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Step 3



Good food is cooked well.

Goal: Analyze data well.

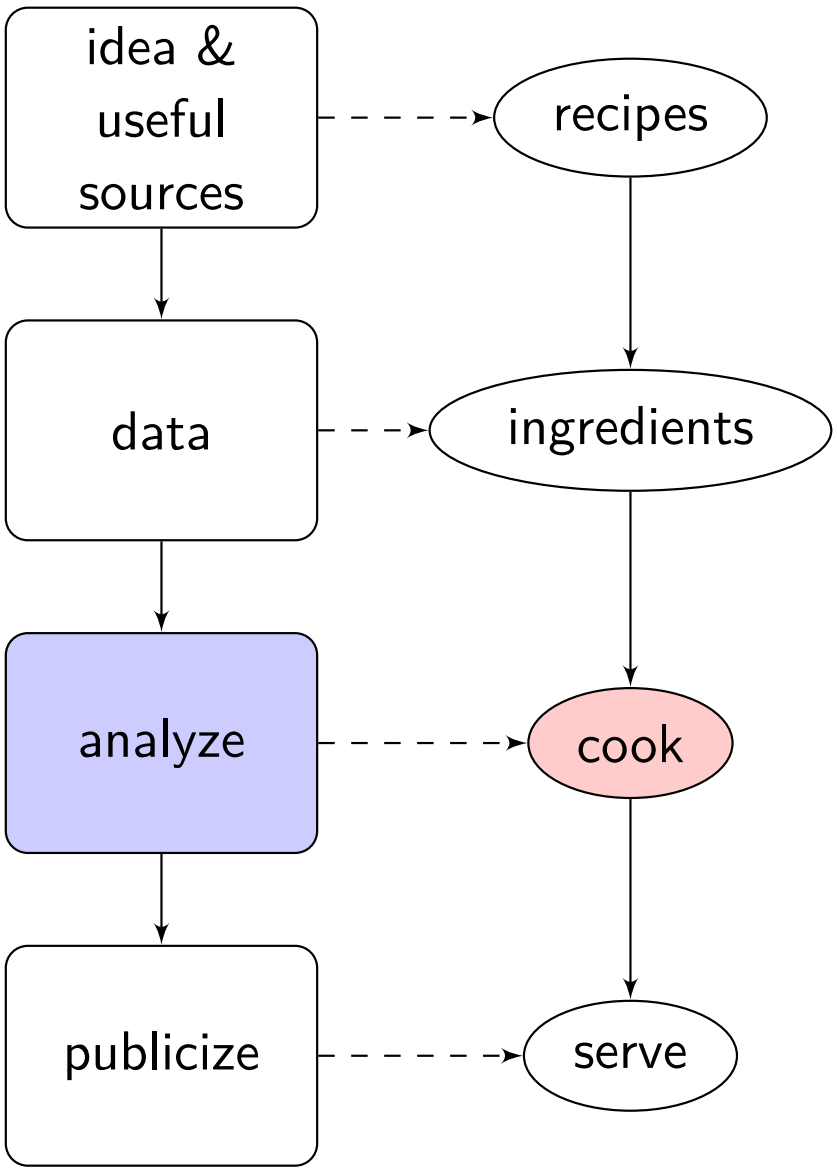
Useful tools for analyzing data:

- Python & R
- (Stata)

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Step 3



Good food is cooked well.

Goal: Analyze data well.

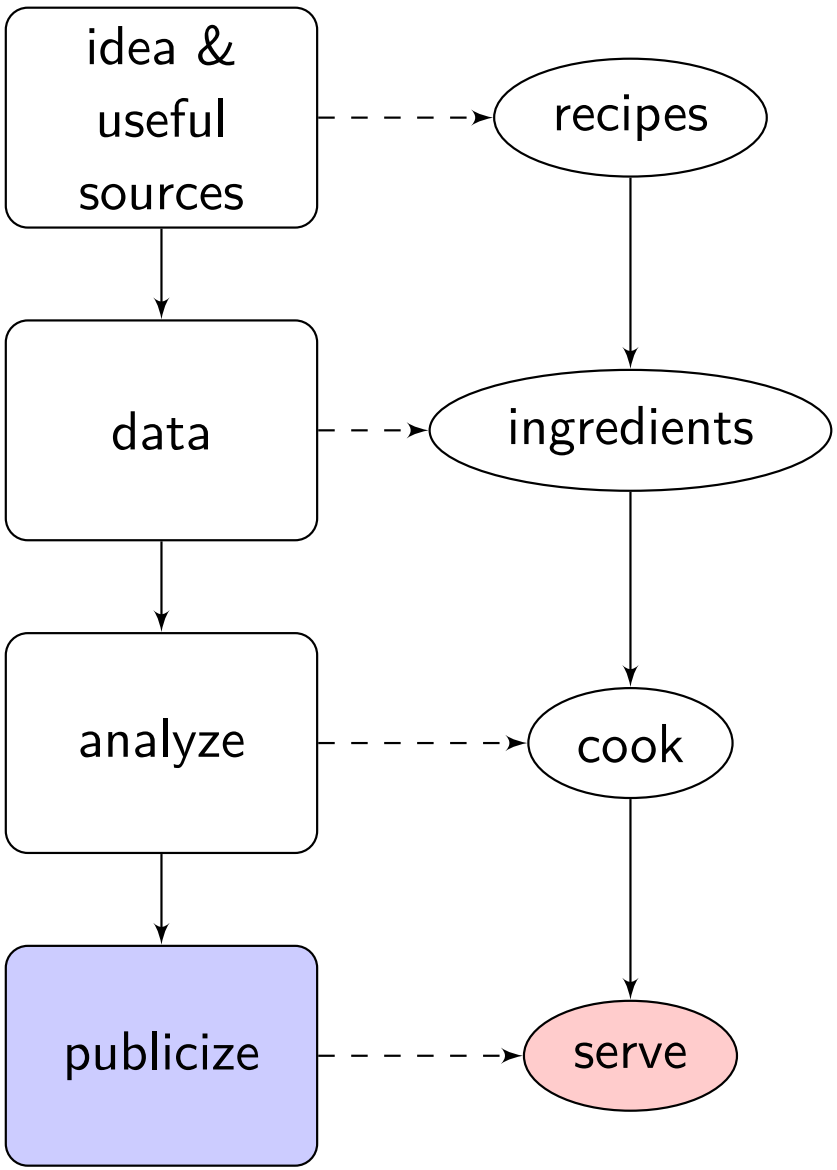
Useful tools for analyzing data:

- Python & R
- (Stata)

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Step 4



Good food is served well.

Goals: Summarize and visualize data well.

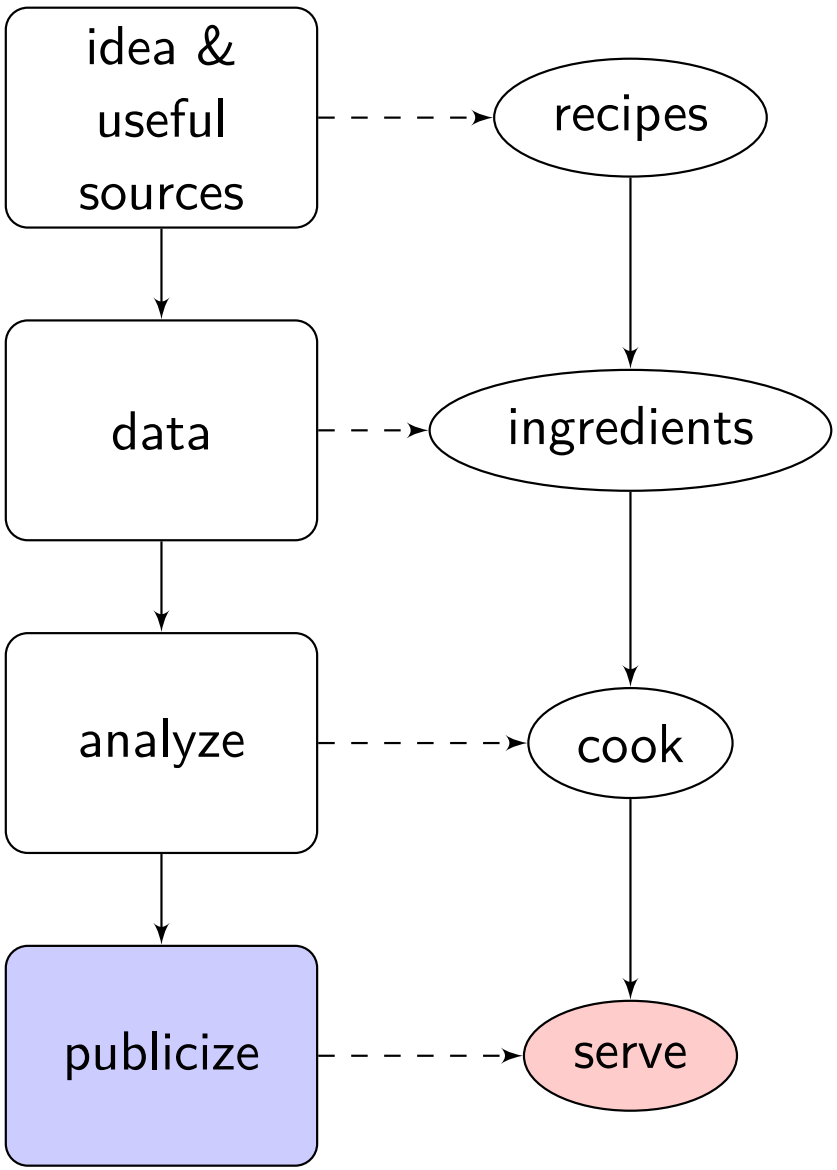
Useful tools for writing a paper and slides:

- LaTeX
- Lyx

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Step 4



Good food is served well.

Goals: Summarize and visualize data well.

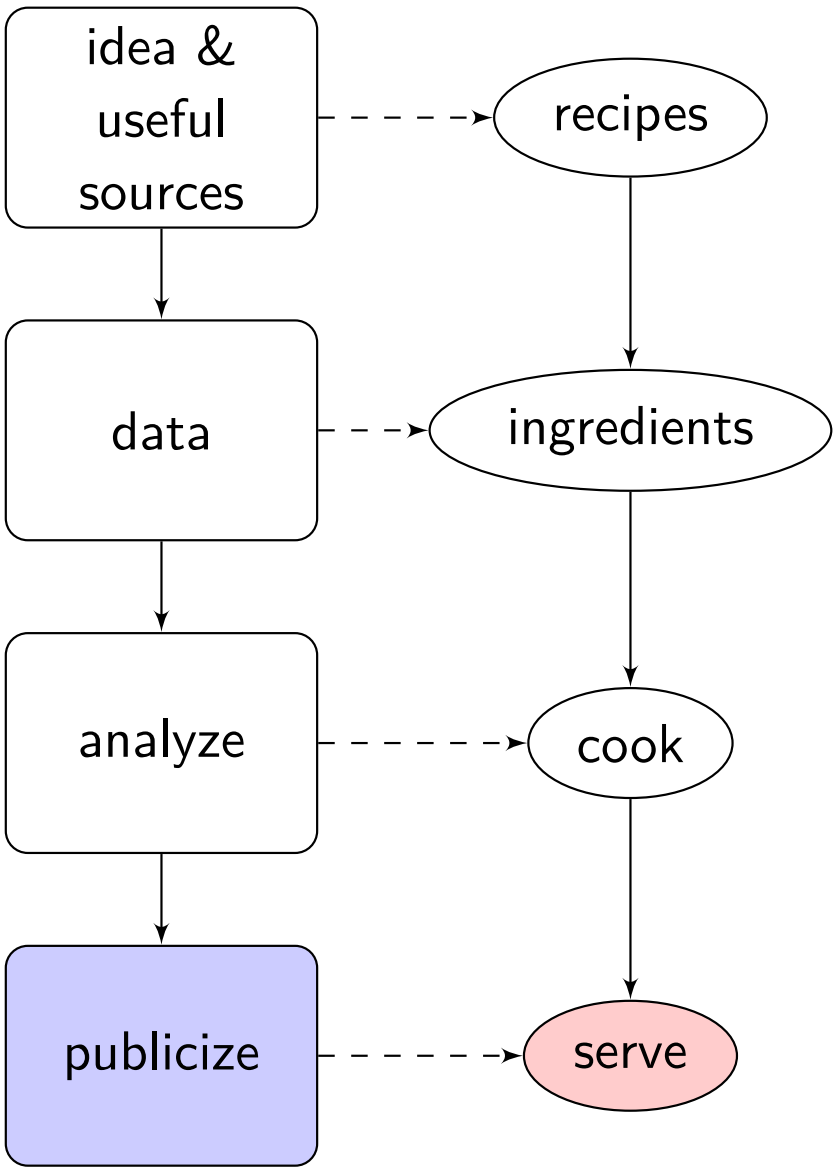
Useful tools for writing a paper and slides:

- LaTeX
- Lyx

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Step 4



Good food is served well.

Goals: Summarize and visualize data well.

Useful tools for writing a paper and slides:

- LaTeX
- Lyx

Useful tools for storing and sharing data and/or code:

- GitHub
- Jupyter Notebook
- Dropbox

Course plan

1. Version control (Git)
2. Python
    - Basics
    - Clean data
    - Analyze data
    - Web scraping

(Assignment 1: Make a panel dataset using Python.)

3. R
    - Basics
    - Clean data
    - Analyze data
    - Text analysis

(Assignment 2: Analyze the panel dataset using R.)

4. Write a paper and slides
5. Work in team

(Term paper)

About the class (cont.)

Each class consists of

- Short lecture, and/or
- Individual/group work with instructions

Idea

Idea $\rightarrow$ very, very important in empirical research

A good idea per 100 mediocre ideas (say)

A good idea is *important* and *feasible*.

- Important = Has a significant contribution in the literature
- Feasible = Possible to test the hypothesis

How to come up with a good idea?

- Read many articles, but don't read so much.

  $\rightarrow$ Just to know topics, and knows and unknowns
- Frequently ask empirical questions to yourself (e.g. while reading news articles).
- Store ideas well and revisit them later.

Read many, but not so much...

The pile of papers I have read during my Ph.D.

Manage idea & resources

Store and organize idea and resources in cloud

- Evernote
- Free
- Unlimited storage
- Web Clipper available for Firefox and Chrome
- Basic account allows sync across only two devices

Why cloud?

- Handy (easy to save/access/organize/share)
- Hard to lose
- Won't fill up local storage

Manage articles

Store and organize articles in cloud

- Readcube
- It's free as long as you use it locally. Online version is free for 30 days, then $55/year.
- Unlimited storage
- Web Importer available for Chrome
- Other options: Mendeley (free), Endnote...

Why cloud?

- Handy (easy to save/access/organize/share)
- Hard to lose
- Won't fill up local storage
- Easy to make reference lists

Manage articles (PDFs) and resources

Store articles (PDFs) and resources in cloud

- Dropbox
- Basic is free. Plus is 99EUR/year.
- 2GB (Basic. If you invite a friend, you will get +16GB.), 1TB (Plus)
- Another option: Google Drive (15GB)

Why cloud?

- Handy (easy to save/access/organize/share)
- Hard to lose

Summary

General workflow

Manage ideas and resources

- Evernote
- Readcube
- Dropbox