

Visit <https://bit.ly/ds-dojo-2024> for material

データ サイエンティスト DOJO

Jonathan Soma

js4571@columbia.edu

Knight Chair in Data Journalism, Columbia University

About me

- Professor in Data Journalism at Columbia Graduate School of Journalism
- Studied Cognitive Science @ UVA
 - Almost computer science...but no!
- Machine Learning and AI @ Columbia
- Worked for New York Times, ProPublica, founded a data news team at the nation's largest radio station
- Non-traditional path into journalism and academia (web development, design, education)



why are you learning
from a data journalist?

it's a very good question.

journalism layoffs...

The screenshot shows a news article from POLITICO. The headline reads: "Over 500 journalists were laid off in January 2024 alone". Below the headline, a sub-headline states: "The job cuts come after an already bleak year in the news business." The article is by Charlotte Tobitt. At the bottom of the page, there is a large image of the Los Angeles Times building with its name prominently displayed on the roof. A caption below the image says: "More than 100 employees were let go from The Los Angeles Times in January. | Patrick T. Fallon/AFP via Getty Images". The author's name, Kierra Frazier, and the date, 02/01/2024 07:31 AM EST, are also visible.

The screenshot shows a news article from Press Gazette. The headline reads: "News media job cuts 2024 tracked: Gannett, Time, Axios, Tampa Bay Times, NYPost, Hollywood Reporter all hit in August". Below the headline, a sub-headline states: "Big losses at the likes of The Messenger, Sports Illustrated and Media Ireland started 2024." The author's name, Charlotte Tobitt, is mentioned. At the bottom of the page, there is a large image of a red CNN logo sign. A caption below the image says: "A giant CNN sign is seen at the entrance to the CNN Center in Atlanta, Georgia, USA, 2022. Picture: Shutterstock/Tada Images".

The screenshot shows a news article from Fast Company. The headline reads: "FC News and media layoffs 2024". The article lists several media companies that have announced layoffs in 2024, including The Los Angeles Times, NBC News, MSNBC, Sports Illustrated, Time, Business Insider, Forbes, TechCrunch, The Messenger, The Wall Street Journal, CBS News, The Intercept, NowThis, BuzzFeed, Vice Media, and WAMU radio. The author's name, Bruce Dixon, is mentioned. The article includes a list of bullet points detailing the layoffs:

- *The Los Angeles Times* laid off 20% of its newsroom in January.
- NBC News and MSNBC laid off around 75 employees in January.
- Sports Illustrated laid off most of its staff (around 100) after it failed to pay licensing fees to its parent company in January.
- Time laid off 15% of its staff, or roughly 30 employees, in January.
- Business Insider CEO Barbara Peng announced a staff reduction of 8% in January.
- Forbes reduced its staff by 3% in late January.
- TechCrunch laid off a handful of staffers and is going to end its paid subscription options.
- The Messenger, a news startup, shut down entirely at the beginning of February after less than a year in operation, leaving more than 300 employees jobless.
- The Wall Street Journal let 20 staff members go at its Washington, D.C., bureau in early February.
- CBS News also cut 20 jobs at its D.C. bureau in early February, as a larger round of 800 cuts at Paramount.
- The Intercept laid off 15 staff members, including its editor-in-chief, in mid-February.
- NowThis cut half of its editorial team in mid-February, a loss of 26 jobs.
- BuzzFeed sold one of its sub-brands, Complex, this week, and subsequently announced a 16% reduction in staff. This comes after shutting its entire news division last year.
- Vice Media will stop publishing on Vice.com and will lay off hundreds, per recent reports.
- WAMU radio, the NPR affiliate in Washington, D.C., said it will shut down the local news website DCist and lay off its staff.

...but data is booming!

Post

Bloomberg Graphics  

@BBGVisualData

Bloomberg is expanding its data journalism and visualization team globally by hiring approximately 40 new data journalists, data visualization reporters, editors and engineers.

a thread 



We're



Hearst Television

Data Journalist

Washington, DC · Reposted 1 week ago · Over 100 applicants



\$95K/yr - \$105K/yr · Full-time · Mid-Senior level



Realtor.com

Data Journalist

Austin, TX · Reposted 1 day ago · 87 applicants



Full-time



hims & hers

Freelance Data Journalist

United States · Reposted 4 days ago · Over 100 applicants



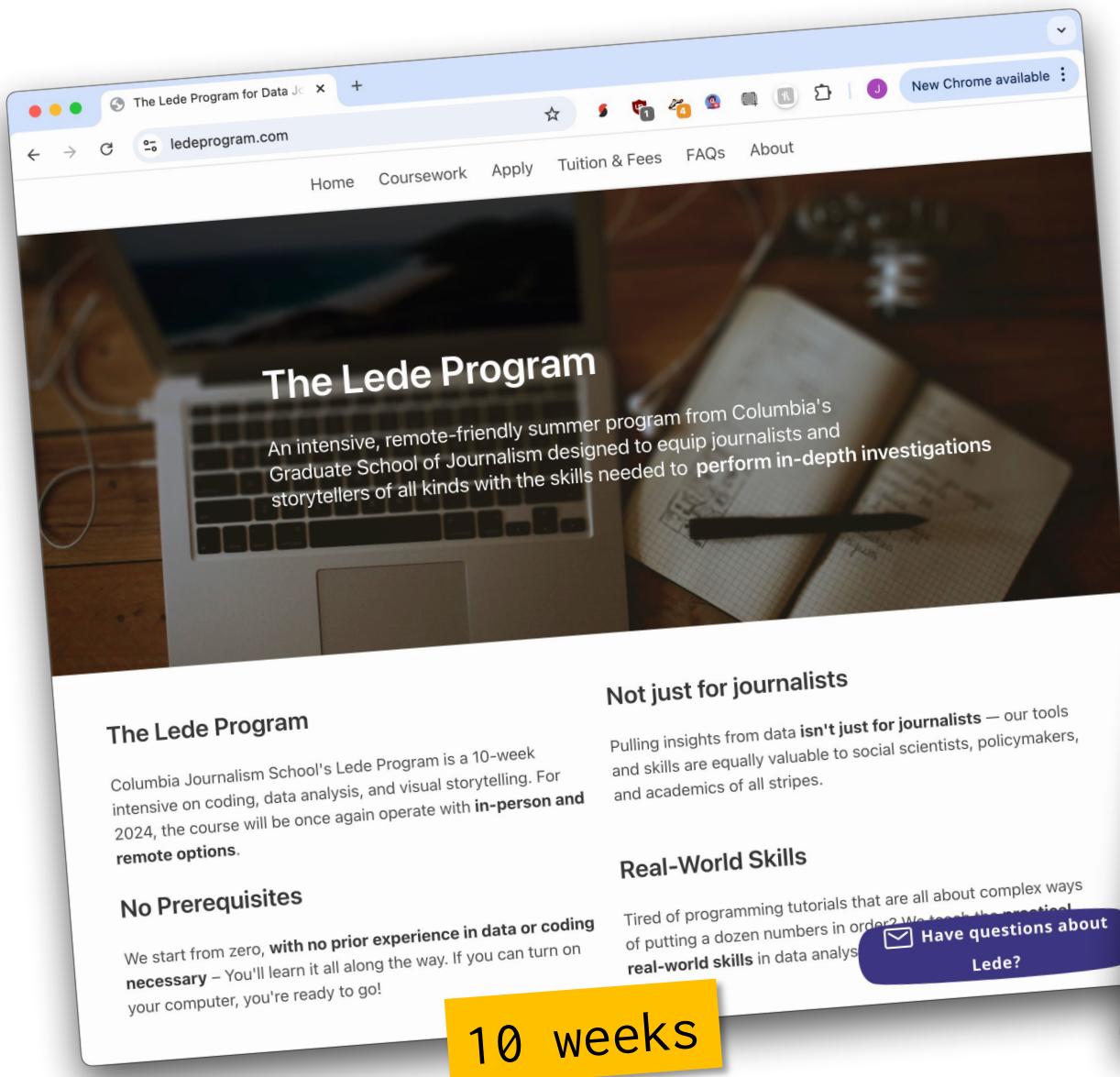
Remote · Full-time · Mid-Senior level

data is flexibility

taking information,
explaining it to others

new business plans

- Non-journalism companies need to communicate
- A journalism organization in Brazil scrapes legislation, organizes and categorizes it, then sells the information to lawyers
- Skilled interactive news and graphics teams do development work for large, well-paying clients to support their journalism work
- **With a little extra knowledge and skills, many many many many more things are possible**



The Lede Program

An intensive, remote-friendly summer program from Columbia's Graduate School of Journalism designed to equip journalists and storytellers of all kinds with the skills needed to perform in-depth investigations

The Lede Program

Columbia Journalism School's Lede Program is a 10-week intensive on coding, data analysis, and visual storytelling. For 2024, the course will be once again operate with **in-person and remote options**.

No Prerequisites

We start from zero, **with no prior experience in data or coding necessary** – You'll learn it all along the way. If you can turn on your computer, you're ready to go!

10 weeks

Not just for journalists

Pulling insights from data **isn't just for journalists** — our tools, and skills are equally valuable to social scientists, policymakers, and academics of all stripes.

Real-World Skills

Tired of programming tutorials that are all about complex ways of putting a dozen numbers in order? We teach the practical of putting a dozen numbers in order? **Have questions about real-world skills in data analysis?**

Lede?



M.S. Data Journalism | Columbia University in the City of New York

COLUMBIA JOURNALISM SCHOOL

Academics ▾ People ▾ Professional Learning ▾ Centers ▾ Community ▾ About ▾

Home Academics M.S. Data Journalism

Program Spotlight: M.S. in Data Journalism

COLUMBIA JOURNALISM SCHOOL

PROGRAM SPOTLIGHT 

M.S. in Data Journalism

Watch on  YouTube

M.S. Data Journalism

Journalism in the 21st century is... **analyzing data for storytelling,**

12 months

データサイエンティストDOJO

nikkeibp.co.jp/seminar/nb/ddojo/?n_cid=emsl_...

特徴 プログラム 開催概要 お問い合わせ お申し込み

米コロンビア大学 ジョナサン・ソーマ教授が来日し直伝

データに裏打ちされた事業計画を立案する力が身につく

マネジメント層とリーダーのための

データサイエンティスト
DOJO(道場)

全6日間集中講座

2024年9月23日(月)開講

米コロンビア大学で開催している
人気プログラムを日本で開催

one week!

事業計画や中期経営計画の立案で、
過去の経験や勘に頼ってはいませんか？

データの裏付けがない計画は、実効力がありません。プロジェクトリーダー自らが、適切な仮説を打ち立て、データの意味を読み解き、深い洞察力を身に着ける必要があります。データサイエンティストとしてのスキルを学ぶにとどまらず、データに裏打ちされた仮説を検証する思考回路を要請するのが、この研修の主たる目的です。「仮説を立てる」→「データを集める」→「データを分析する」→「ビジュアライゼーションによって、理解と洞察力を深化する」→「チームで共有し、次なる仮説を立てる」このサイクルを繰り返すことによって、より精度の高い事業計画や中期経営計画の立案が可能になります。

Goals

- We are not aiming for expertise!
- We are aiming for foundational knowledge
- We want to be adaptable

“with time, I can
accomplish this”

...almost? maybe?

What you don't know, you can learn.
With a good foundation, everything is
possible.

introductions

Turn to the people next to you, in front of you, behind you. Introduce yourself.
Where are you from + why are you here?

ask questions!!

important!!!!

asking questions is **how you learn**, and I am sure everyone wants to know the same thing

Our schedule

- **Monday through Friday + Monday**
 - Class during day
- **Homework**
 - Optional!
 - A large amount for learning later
- **This Friday:** Working on personal projects
- **Next Monday:** Project presentations



What we'll learn

- **Coding:** Fundamentals of Python
- **Data analysis:** pandas (Excel for Python)
- **Structured data:** APIs
- **Unstructured data:** Scraping
- **AI tricks:** Transcription, data tools, app building
- **Data visualization:** Coding and no-code tools
- **Project planning:** Make sure you finish!

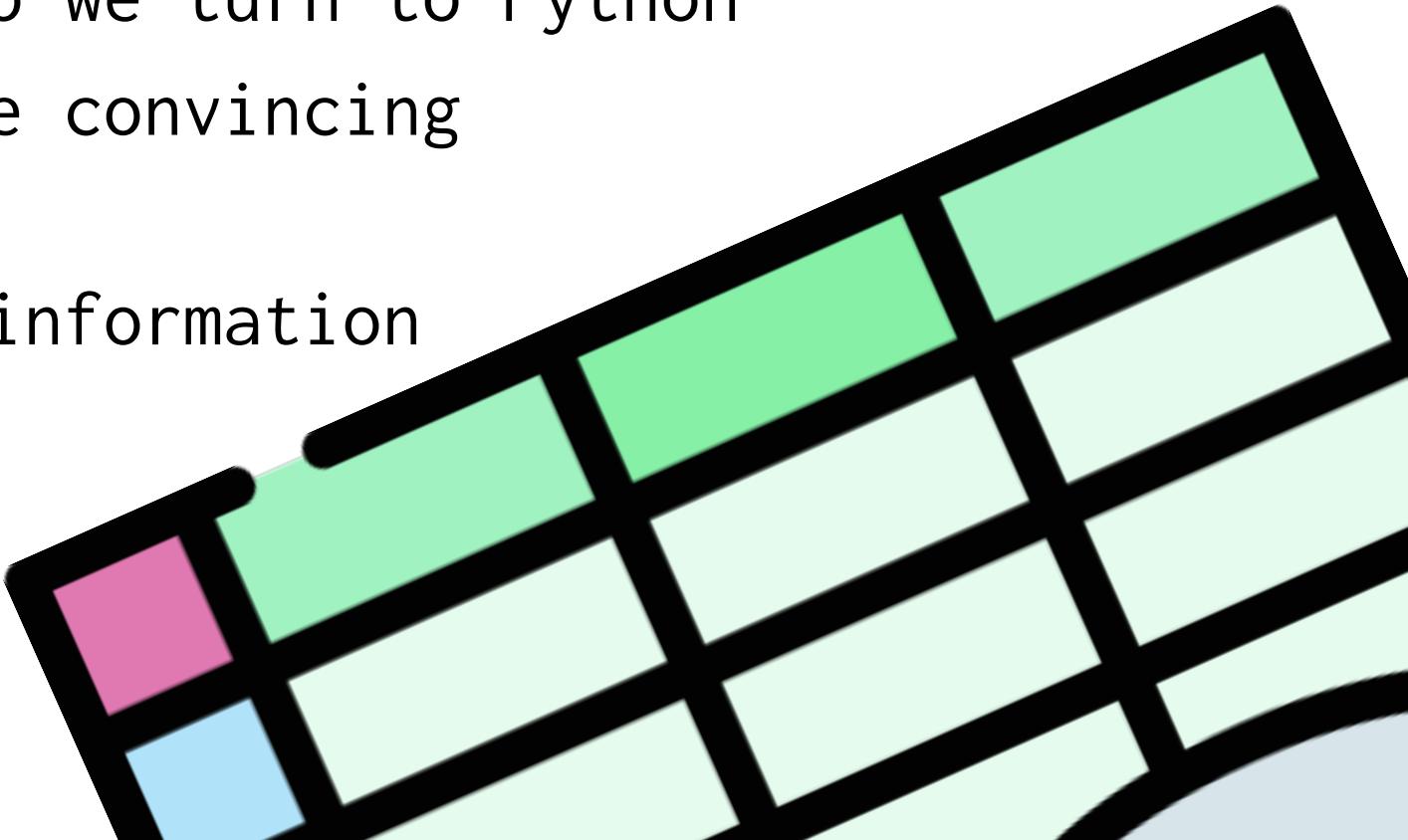
Python programming

- Fundamental to every other skill
- A superpower that allows you to do anything
- Not just “coding”: **Python is special**
 - Academic for research
 - Industry for software development
 - Online tools, desktop applications, databases
 - Powers most of the AI tools



Data analysis

- A person produces ~6GB of data per hour
- Excel has limitations, so we turn to Python
- Data makes arguments more convincing
 - ...but that makes me sad
- Just one more source of information

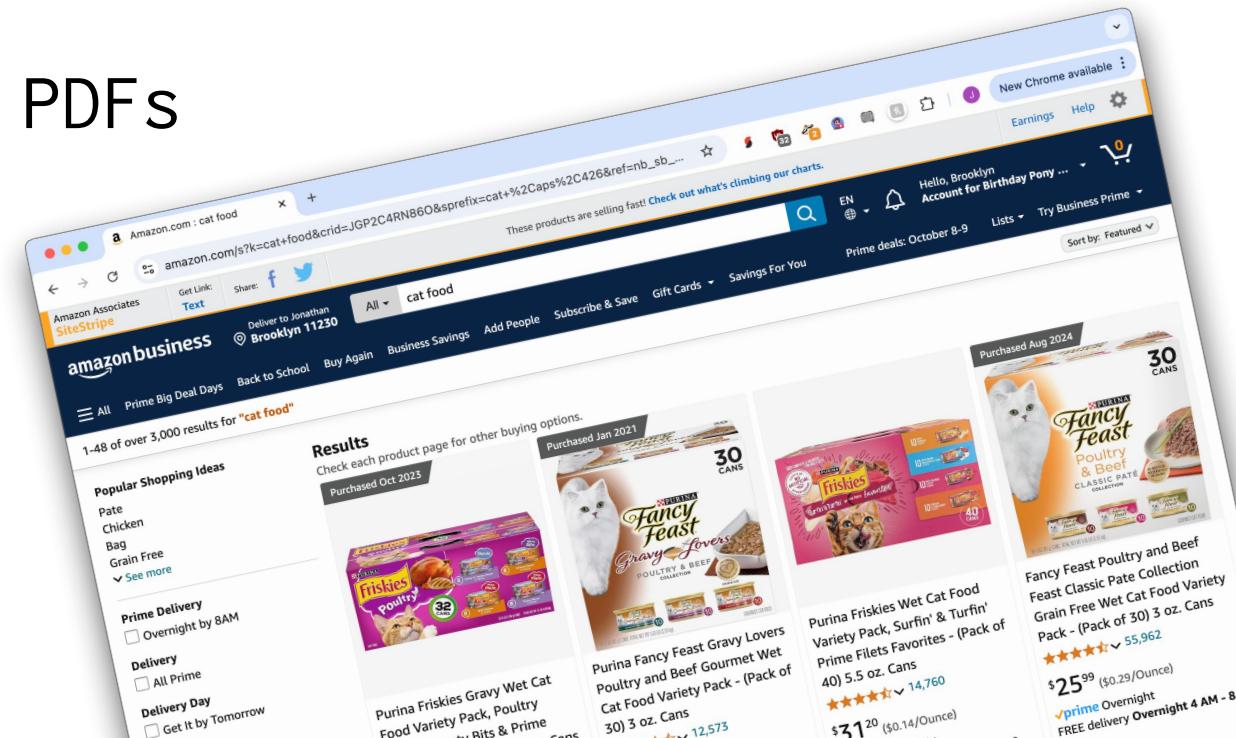


Structured data

- Data that is easy for computers to understand: Excel, CSV, XML, JSON, APIs, databases
 - **APIs:** Computers talking to computers
 - Available from governments and industry
 - Easy to reformat or process to display to members of your company or customers
 - Add value by combining with other data

Unstructured data

- Data that is easy for **people** to understand
 - Our job is to make computers understand it
- Scraping data from websites
- Extracting data from text or PDFs
- Add value by processing it!



AI tricks

- Accomplish impossible tasks!
- Be successful with skills you **don't have**
- Get better with skills you **do have**
- Automate “grunt work” and “intern jobs”
- It's for more than rewriting emails :)

Data visualization

- Communicating with data is difficult
- Sell your content more effectively
- Everyone loves charts and graphs!

Project Planning

- Estimating time is very difficult for tech projects
- Finishing is the hardest part

Hello world: AI !

we have a secret weapon,
it's artificial intelligence

The background of the image features a dense, repeating floral and foliate pattern in shades of blue, yellow, and pink, set against a dark, solid-colored backdrop. The pattern includes large stylized flowers, smaller blossoms, and delicate leaves.

A.I. is not

perfect

A.I. should be used for

error-resilient tasks

JS

You

Write a sonnet about my fat, mean cat Mulberry.



ChatGPT

Beneath the moon, Mulberry takes her throne,
 A cat of girth, with fur like midnight's shade.
 Her eyes, twin emeralds, cold as stone,
 Within their depths, her fierce commands are made.

Her claws, like daggers, glint in soft moonlight,
 A queen who rules with strict and stern decree.
 Her growl can turn the bravest heart to fright.
 Her hiss a symphony of sharp, silken strings.

low-quality output is ignorable

when shadows dance and silent

She curls beside me, softening I

Revealing tender moments, warr

For though she's fat and often se

In dreams, sweet Mulberry's a gei

```
# Python function to calculate fibonacci sequence
def fibonacci(n):
    if n == 0:
        return 0
    elif n == 1:
        return 1
    else:
        return fibonacci(n-1) + fibonacci(n-2)
```

tight feedback loops

errors are expected

New York Sessions

Year **Session**

2023-
2024

General Assembly

20091

3380

[Browse](#)

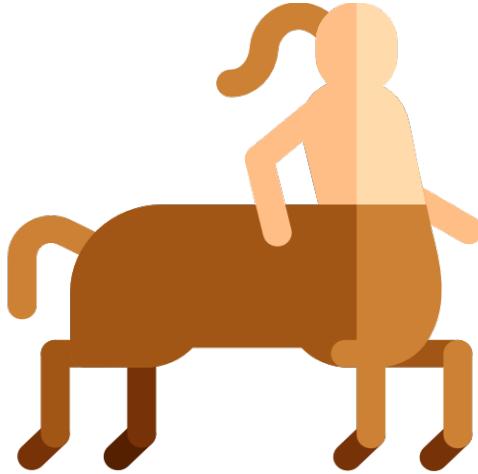
(a.k.a. "intern work")



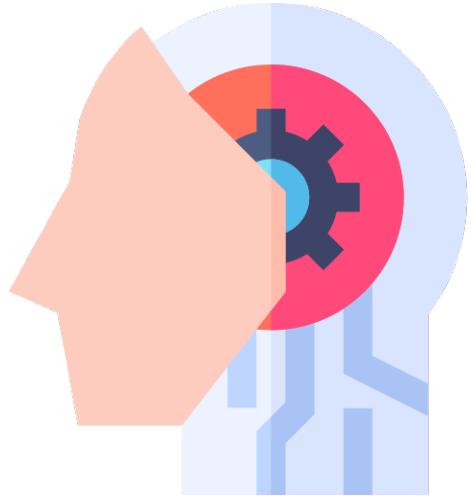


```
# Python function to calculate fibonacci sequence
def fibonacci(n):
    if n == 0:
        return 0
    elif n == 1:
        return 1
    else:
        return fibonacci(n-1) + fibonacci(n-2)
```

tight feedback loops



Centaurs and cyborgs



- Clear division of labor between AI and person
- One task for AI, one task for human
- **Intuition** of humans, **trust** in AI
- Intertwined, deep integration of AI into human work
- **Collaborative**, back-and-forth process

we will be both!

that will be our week!

any questions?

Time for a survey!



<https://bit.ly/ds-dojo-survey>

Software Installation

Usually the worst part.
This time, not terrible!

Running+Installing Python

The image displays two screenshots of software interfaces for managing Python installations.

Python Wrangler: A dark-themed application window titled "Python Wrangler". It features a sidebar with links for "Py", "Python Wrangler", "Pythons", "Walkthrough", and "About". On the right, a table titled "Python Installations" lists 19 installed Python environments. The columns are "Command", "Version", "Installation", "Packages", and "Location(s)".

Command	Version	Installation	Packages	Location(s)
python	3.6.4	pyenv 2018-04-15	75 packages	/Users/jonathansoma/.pyenv/shims/python in PATH
python3	3.6.4	pythonorg 2018-02-25	18 packages	/Library/Frameworks/Python.framework/Versions/3.6/bin/python3.6 /Library/Frameworks/Python.framework/Versions/3.6/bin/python3 in PATH symlink /usr/local/bin/python3 in PATH symlink
	3.6.1	homebrew 2017-03-23	200 packages	/usr/local/Cellar/python3/3.6.1/Frameworks/Python.framework/Versions/3.6/bin/python3 /usr/local/Cellar/python3/3.6.1/bin/python3 symlink
	2.7.13	homebrew 2017-04-21	130 packages	/usr/local/Cellar/python/2.7.13/Frameworks/Python.framework/Versions/2.7/bin/python /usr/local/Cellar/python/2.7.13/bin/python symlink
	2.7.13	homebrew 2017-04-21	130 packages	/usr/local/Cellar/python3/2.7.13/Frameworks/Python.framework/Versions/2.7/bin/python3 /usr/local/Cellar/python3/2.7.13/bin/python3 symlink
	3.6.4	homebrew 2018-02-15	200 packages	/usr/local/Cellar/python/3.6.4_4/Frameworks/Python.framework/Versions/3.6/bin/python3 /usr/local/Cellar/python/3.6.4_4/bin/python3 symlink

Anaconda Navigator: A light-themed application window titled "ANACONDA NAVIGATOR". It has tabs for "Home", "Documentation (26)", "Training (0)", and "Connect". The main area shows a list of installed Python environments, which is identical to the one in Python Wrangler.

Running+Installing Python

- **Anaconda:** Popular with data scientists
 - Simple installation
- **Visual Studio Code:** Popular for web site development
 - Integration with many common tools
- **Google Colab:** In the cloud, popular for AI development
 - Easy access to powerful computer resources
- **Jupyter:** Popular with data sciences and data journos
 - Good for having “conversations” with your data
- **It doesn’t matter which you use!!!!!!** They are all a little difficult to set up and use.

JupyterLab Desktop

<https://bit.ly/ds-jupyterlab>

- Windows, Macs, Linux
- Very easy to install
- **No trouble with Python**
- Easy to change to more “professional” tools later



A screenshot of a web browser showing the GitHub repository for JupyterLab Desktop. The URL is github.com/jupyterlab/jupyterlab-desktop. The repository page displays various sections like Code, Issues, Pull requests, Actions, Projects, Security, and Insights. A large yellow box with the number 1 contains the text "Visit <https://bit.ly/ds-jupyterlab>". A red box with the number 2 contains the text "Scroll down". A large red arrow points downwards from the "Scroll down" text towards the bottom of the repository page, indicating where to scroll.

A screenshot of a web browser showing the GitHub repository for JupyterLab Desktop. The URL is github.com/jupyterlab/jupyterlab-desktop. The page shows the README, Code of conduct, BSD-3-Clause license, and Security sections. A large yellow box with the number 1 contains the text "Visit <https://bit.ly/ds-jupyterlab>". A red box with the number 2 contains the text "Scroll down". A large red arrow points upwards from the "Installation" section towards the top of the page, indicating where to scroll. The "Installation" section provides links for Windows (10, 11), Mac (macOS 10.15+), and Linux, including x64 Installer, arm64 Installer (Apple silicon), Snap Store [recommended], x64 Installer (Intel chip), .deb x64 Installer (Debian, Ubuntu), and .rpm x64 Installer (Red Hat, Fedora, SUSE). It also mentions winget installation on Windows.

Which one to pick?

Windows PC

Installation

Install JupyterLab Desktop using one of the methods listed below for your system.

Windows (10, 11)	Mac (macOS 10.15+)	Linux
x64 Installer	arm64 Installer (Apple silicon) x64 Installer (Intel chip)	Snap Store [recommended]
		.deb x64 Installer (Debian, Ubuntu) .rpm x64 Installer (Red Hat, Fedora, SUSE)

Older macs, not M1/M2/m3

Newer M1 , M2 , M3 macs

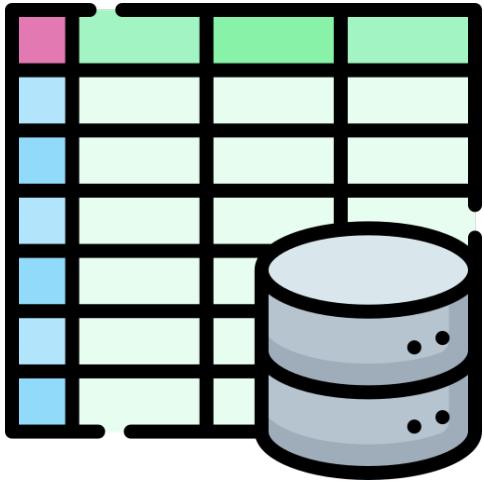
iPads or Chromebooks:
You'll use [Google Colab](#)

Installation

<https://bit.ly/ds-jupyterlab>

- Download and run the installer
 - **Security warning?** Say yes!
- After installation, open the **JupyterLab Desktop** app
- **Maybe it maybe will want to install Python:** there will be a banner to click on the bottom.
- Grey “New notebook?” Close, open again, look for Python install





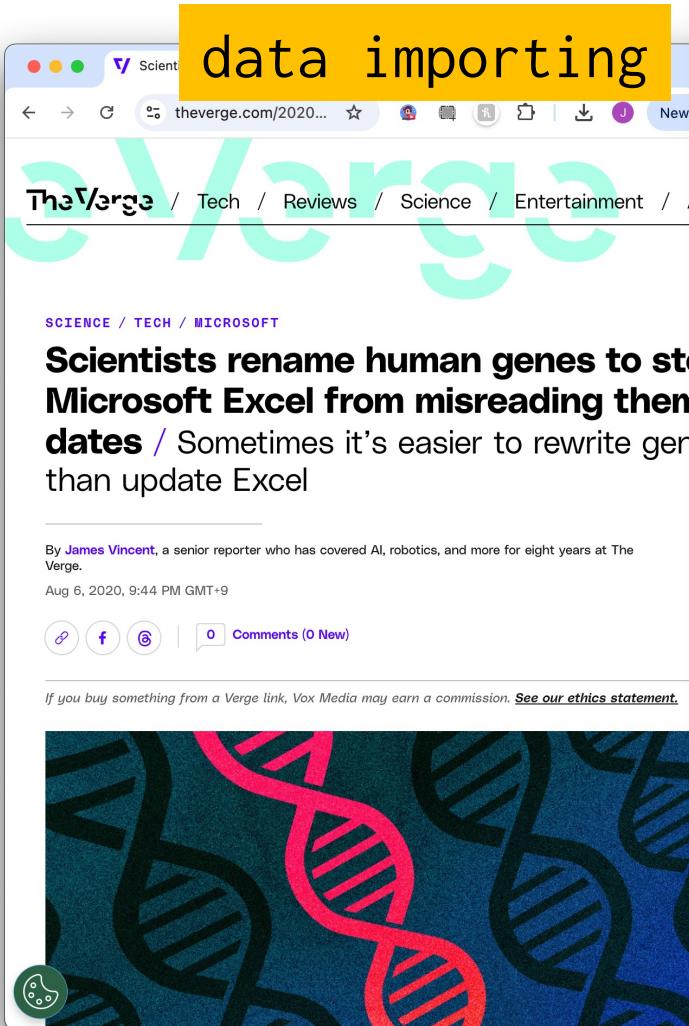
Excel vs. Python



- Easy to get started
- Easy to see results
- Common in business environments
- **Easy to make mistakes**
- Difficult to install
- Difficult to begin learning
- Difficult to share with non-programmers
- **Everything is possible!**

Problems with Excel

data importing



The Verge / Tech / Reviews / Science / Entertainment / AI / SCIENCE / TECH / MICROSOFT

Scientists rename human genes to stop Microsoft Excel from misreading them as dates

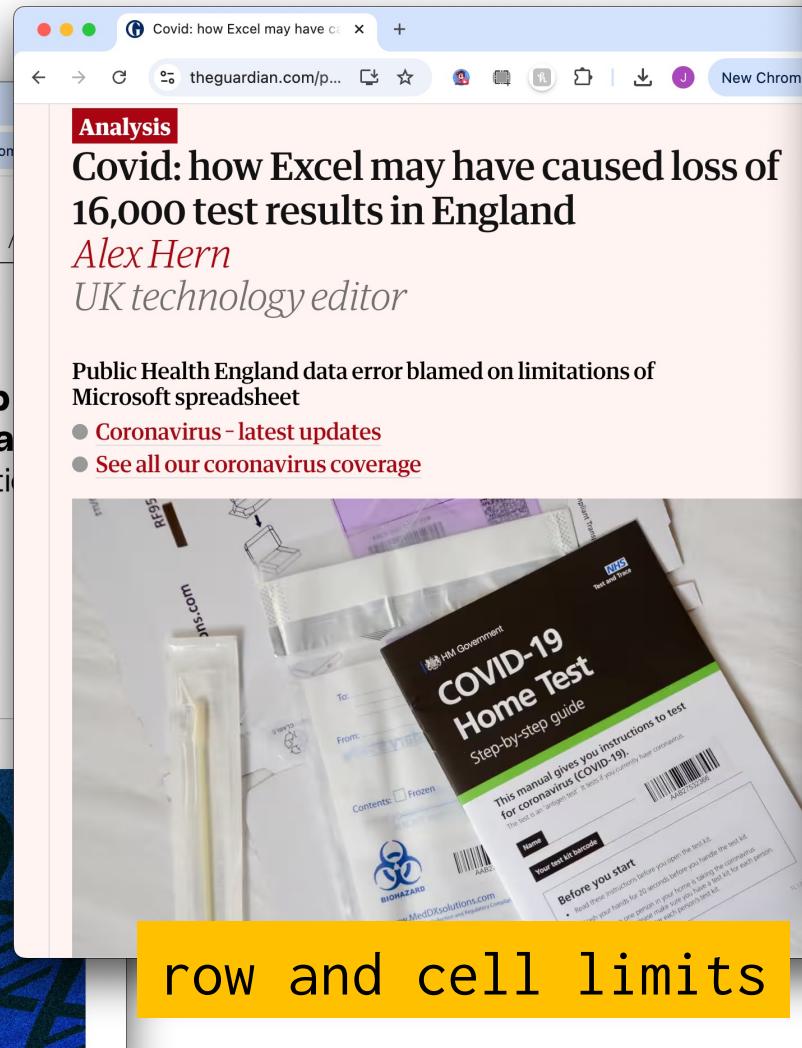
Sometimes it's easier to rewrite genetic than update Excel

By James Vincent, a senior reporter who has covered AI, robotics, and more for eight years at The Verge.

Aug 6, 2020, 9:44 PM GMT+9

0 Comments (0 New)

If you buy something from a Verge link, Vox Media may earn a commission. [See our ethics statement.](#)



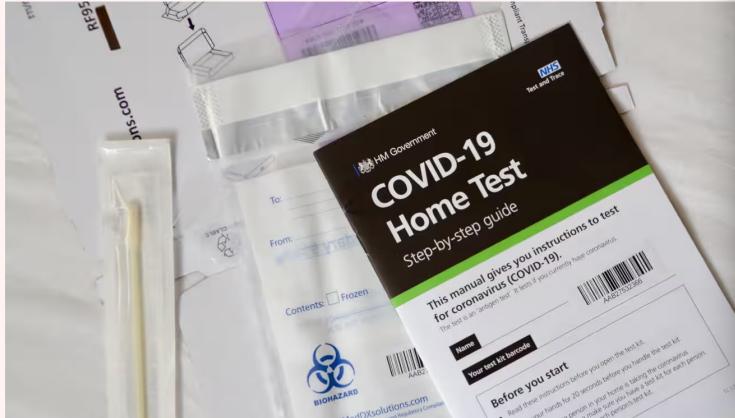
Analysis

Covid: how Excel may have caused loss of 16,000 test results in England

Alex Hern
UK technology editor

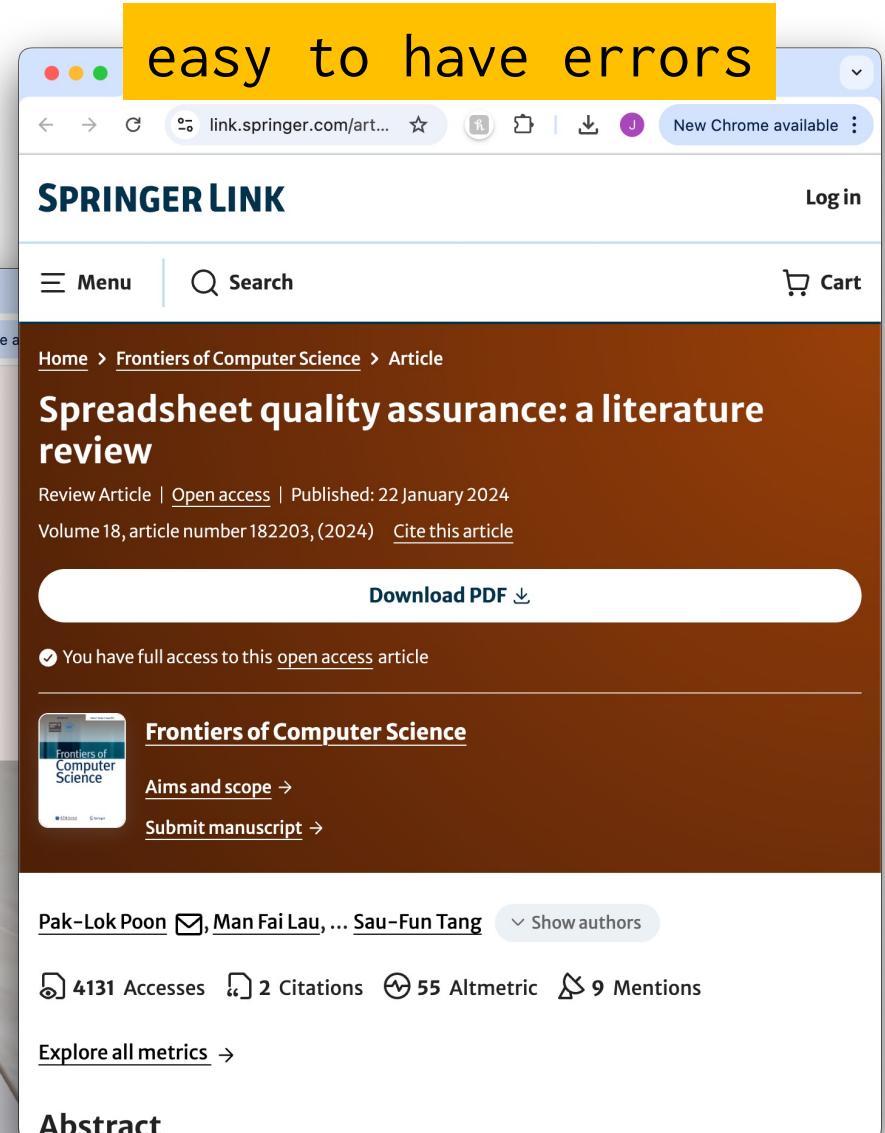
Public Health England data error blamed on limitations of Microsoft spreadsheet

- Coronavirus - latest updates
- See all our coronavirus coverage



row and cell limits

easy to have errors



SPRINGER LINK

Menu Search Log in Cart

Home > Frontiers of Computer Science > Article

Spreadsheet quality assurance: a literature review

Review Article | Open access | Published: 22 January 2024

Volume 18, article number 182203, (2024) Cite this article

Download PDF

You have full access to this open access article

Frontiers of Computer Science

Aims and scope →

Submit manuscript →

Pak-Lok Poon, Man Fai Lau, ... Sau-Fun Tang Show authors

4131 Accesses 2 Citations 55 Altmetric 9 Mentions

Explore all metrics →

Abstract

Excel errors

94% have faults!

Along with the high popularity of spreadsheet applications [10], it was found that about 94% of the spreadsheets in use contained faults [11,12]. A major reason for a high number of faulty spreadsheets is the accelerating trend in end-user computing (or end-user programming) over the last few decades [13–15]. Spreadsheet development, now a prominent example of end-user computing [16,17], has shifted from being often done by well-trained IT professionals to something millions of non-technical departmental end users or *end-user programmers* are now responsible to do. As most end-user programmers are not well trained in software development and testing [18], it is not surprising that many spreadsheets they developed are poorly coded and inadequately tested [19]. Consequently, these spreadsheets are likely to contain faults that are not properly detected and removed before release for daily operational use.

You can read Python

What is happening?

In [27]:

```
df['pct_15-64'] = df['pop_15-64'] / df['total']
df['pct_65-over'] = df['pop_65-over'] / df['total']
df.head()
```

Out[27]:

	YEAR	AREA	/ITEMS	pop_15-64	pop_65-over	total	pct_15-64	pct_65-over
141	2019	Hokkaido	NaN	1012000	1673000	4685000	0.642903	0.357097
142	2019	Iwate-ken	NaN	699000	415000	1114000	0.627469	0.372531
143	2019	Miyagi-ken	NaN	684000	406000	1090000	0.627523	0.372477
144	2019	Akita-ken	NaN	1382000	652000	2034000	0.679442	0.320558

Easy to read

A1030	B	C	D	E
869 長野県 富士見町	203629	1,053	1,105	1,311
870 長野県 原村	203637	1,29	1,273	1,173
871 長野県 辰野町	203823	1,151	1,223	1,424
	203831	1,083	1,165	1,366
				1,269
				1,451
				1,223
				1,399
				1,268

Visit <https://bit.ly/ds-dojo-2024> for material

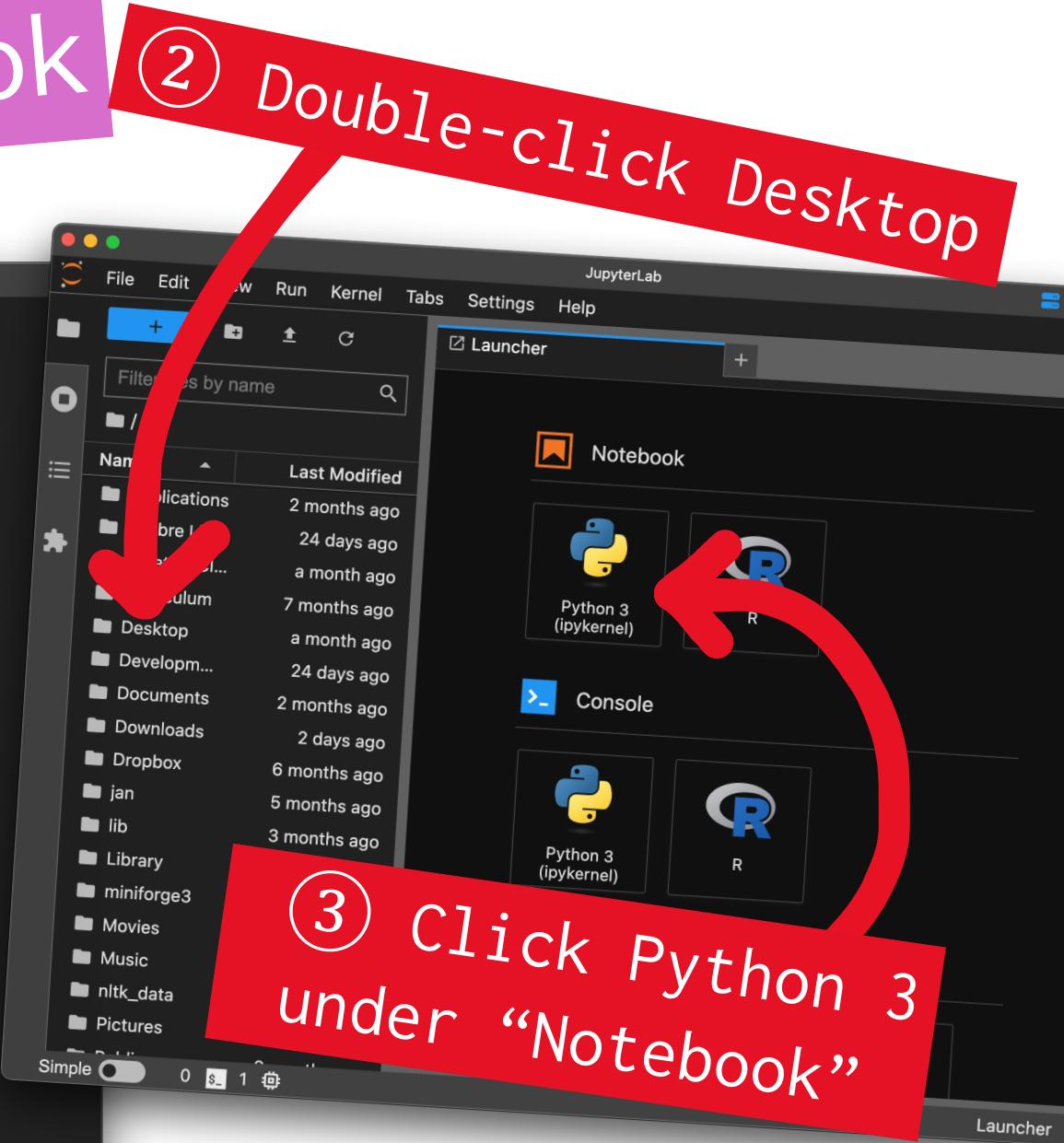
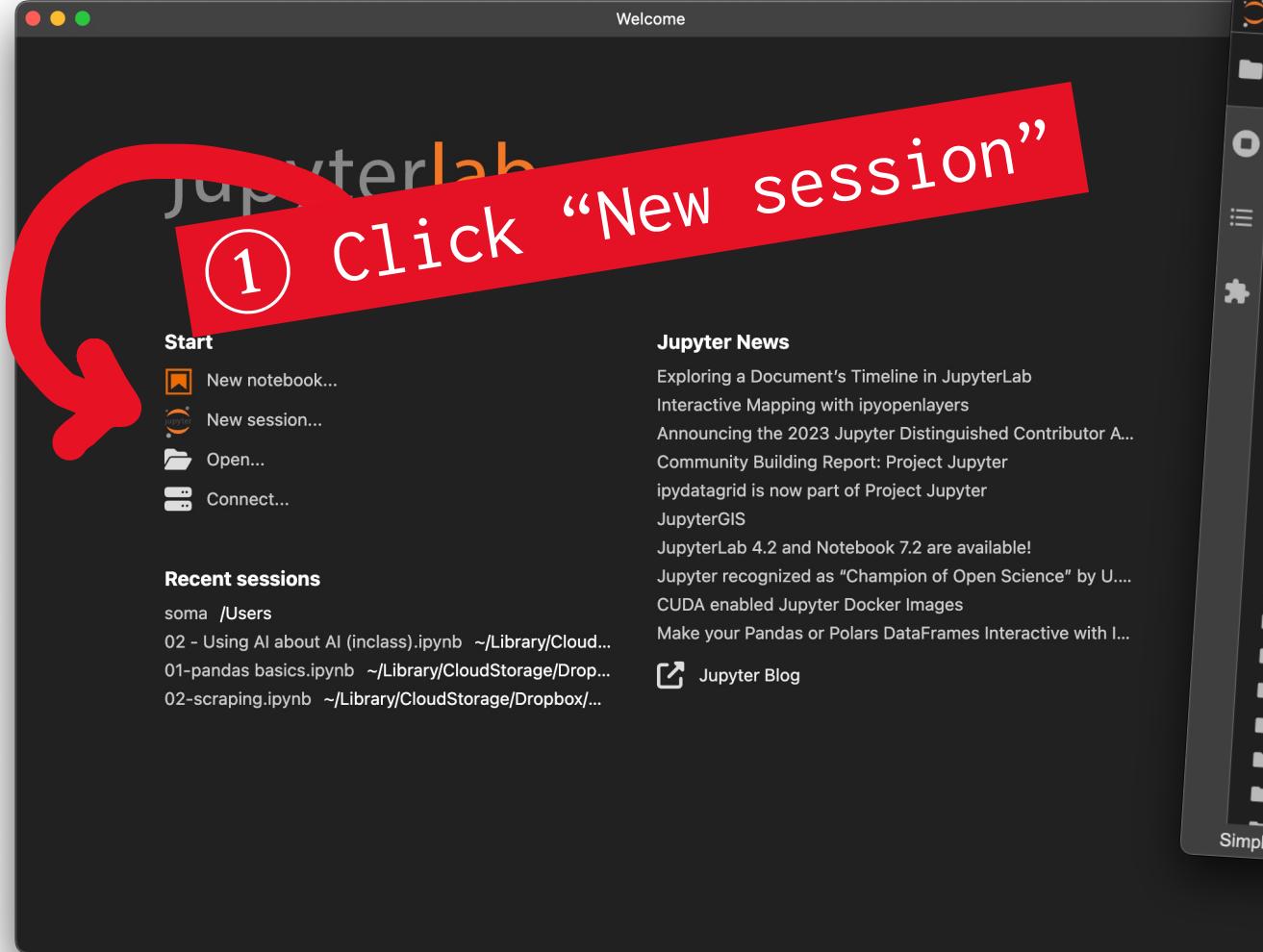
データ サイエンティスト DOJO

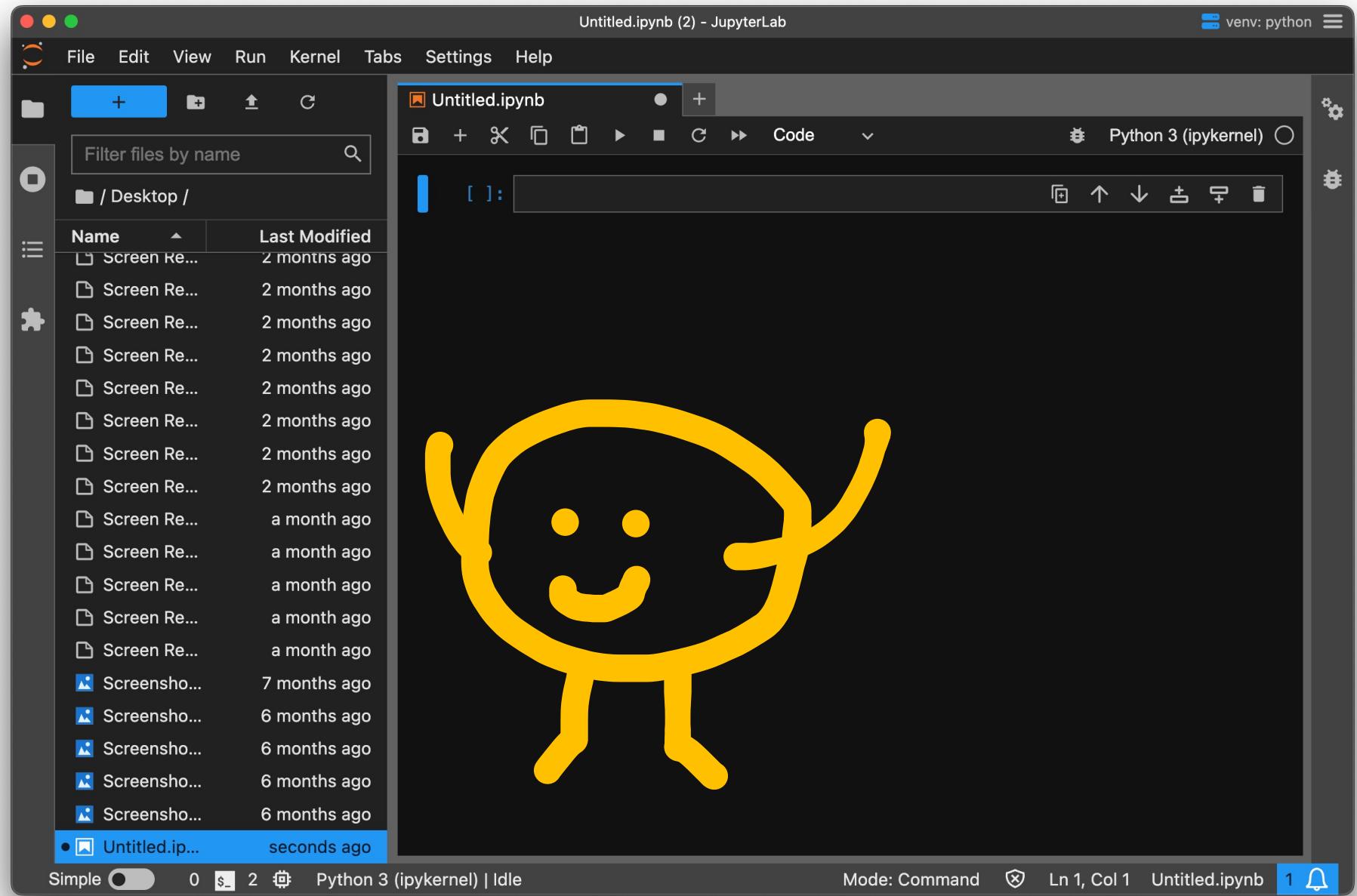
Jonathan Soma

js4571@columbia.edu

Knight Chair in Data Journalism, Columbia University

Create a new notebook





Untitled.ipynb

[]: `print("Hello world")`

Python 3 (ipykernel)

Press Play or Shift+Enter/Return



You should see this

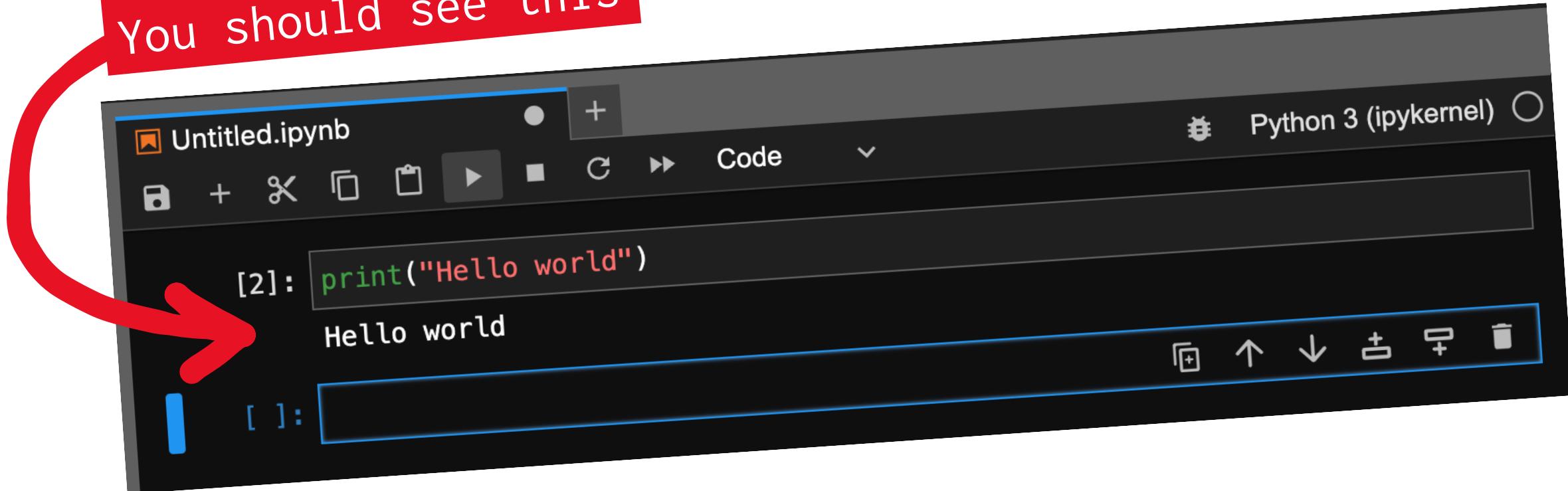
Untitled.ipynb

[2]: `print("Hello world")`

Hello world

[]:

Python 3 (ipykernel)



AutoSave ⚡ ⌂ ⌄ ⌅ ⌆ ⌇ ⌈ ⌉ ⌊ ⌋

Home Insert Draw



Character encoding

Cell Formatting

Table

Cell Styles

Insert

Delete

Format

Σ

Filter

Sort & Filter

Find & Select

Add-ins

Analyze Data

Create PDF

and share link

Comments

Share

Save As...

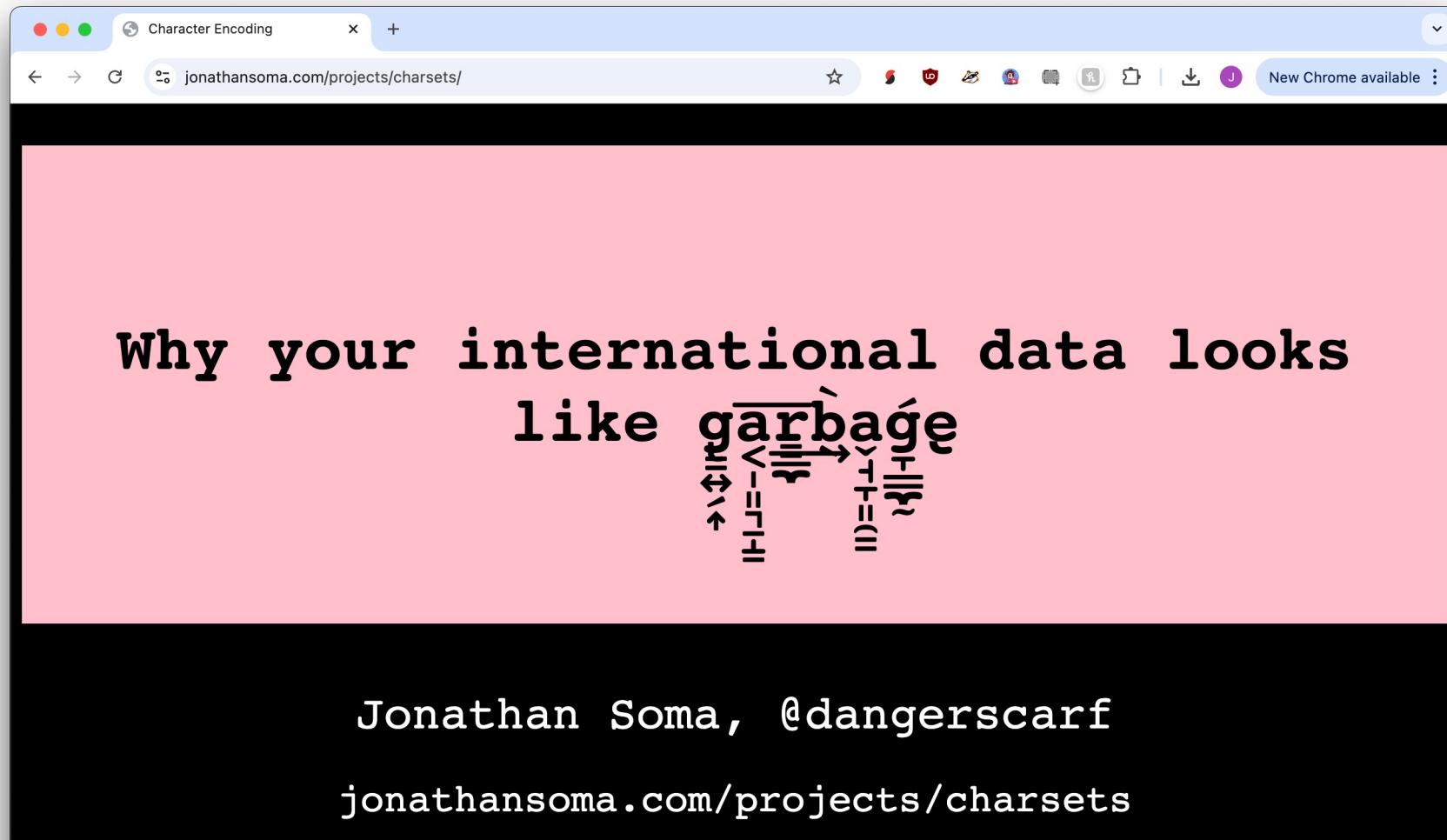
	A	B	C	D	E	F	G
1	name	price	tax	image_url	tags		
2	CORN	298	¥298 (8%tax)	https://www.national-azabu.net/upload/save_image/0607133157_629ed4bd3cb1b.jpeg	PERISHABLE 冷藏		
3	YUZU CITRON	358	¥358 (8%tax)	https://www.national-azabu.net/upload/save_image/1227165628_5fe83e2c165a4.jpeg	PERISHABLE 冷藏		
4	BROCCOLI	398	¥398 (8%tax)	https://www.national-azabu.net/upload/save_image/18_s.jpg	PERISHABLE 冷藏		
5	CUCUMBER 1PC	128	¥128 (8%tax)	https://www.national-azabu.net/upload/save_image/230051000000_s.jpg	PERISHABLE 冷藏		
6	CUCUMBERS 4PCS	498	¥498 (8%tax)	https://www.national-azabu.net/upload/save_image/230034000000_s.jpg	PERISHABLE 冷藏		
		138	¥138 (8%tax)	https://www.national-azabu.net/upload/save_image/2_s.jpg	PERISHABLE 冷藏		
		398	¥398 (8%tax)	https://www.national-azabu.net/upload/save_image/230034000000_s.jpg	PERISHABLE 冷藏		
		398	¥398 (8%tax)	https://www.national-azabu.net/upload/save_image/16_s.jpg	PERISHABLE 冷藏		

products_output.csv



Open with Numbers

name	price	tax	image_url	tags
CORN	298	¥298 (8%tax)	https://www.national-azabu.net/upload/save_image/0607133157_629ed4bd3cb1b.jpeg	PERISHABLE 冷藏
YUZU CITRON	358	¥358 (8%tax)	https://www.national-azabu.net/upload/save_image/1227165628_5fe83e2c165a4.jpeg	PERISHABLE 冷藏
BROCCOLI	398	¥398 (8%tax)	https://www.national-azabu.net/upload/save_image/18_s.jpg	PERISHABLE 冷藏
CUCUMBER 1PC	128	¥128 (8%tax)	https://www.national-azabu.net/upload/save_image/230051000000_s.jpg	PERISHABLE 冷藏
CUCUMBERS 4PCS	498	¥498 (8%tax)	https://www.national-azabu.net/upload/save_image/230034000000_s.jpg	PERISHABLE 冷藏
CARROT 1PC	138	¥138 (8%tax)	https://www.national-azabu.net/upload/save_image/2_s.jpg	PERISHABLE 冷藏
CARROTS 3PCS	398	¥398 (8%tax)	https://www.national-azabu.net/upload/save_image/230034000000_s.jpg	PERISHABLE 冷藏
ORGANIC CARROTS BIG PACK 1KG	398	¥398 (8%tax)	https://www.national-azabu.net/upload/save_image/16_s.jpg	PERISHABLE 冷藏



<https://jonathansoma.com/projects/charsets/>

Python basics

I hope it is review, but if
not: that's okay!

A tiny Python quiz

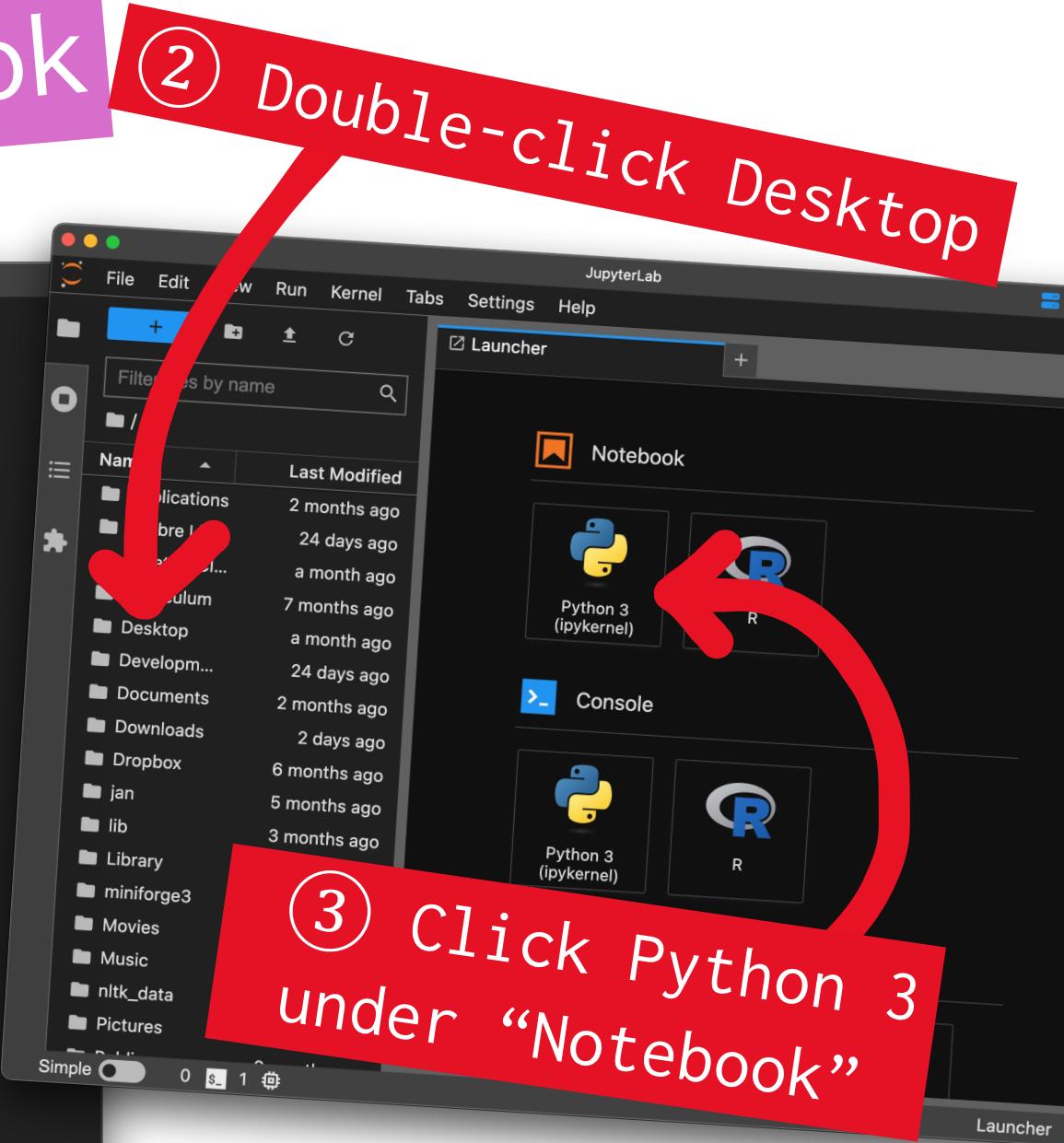
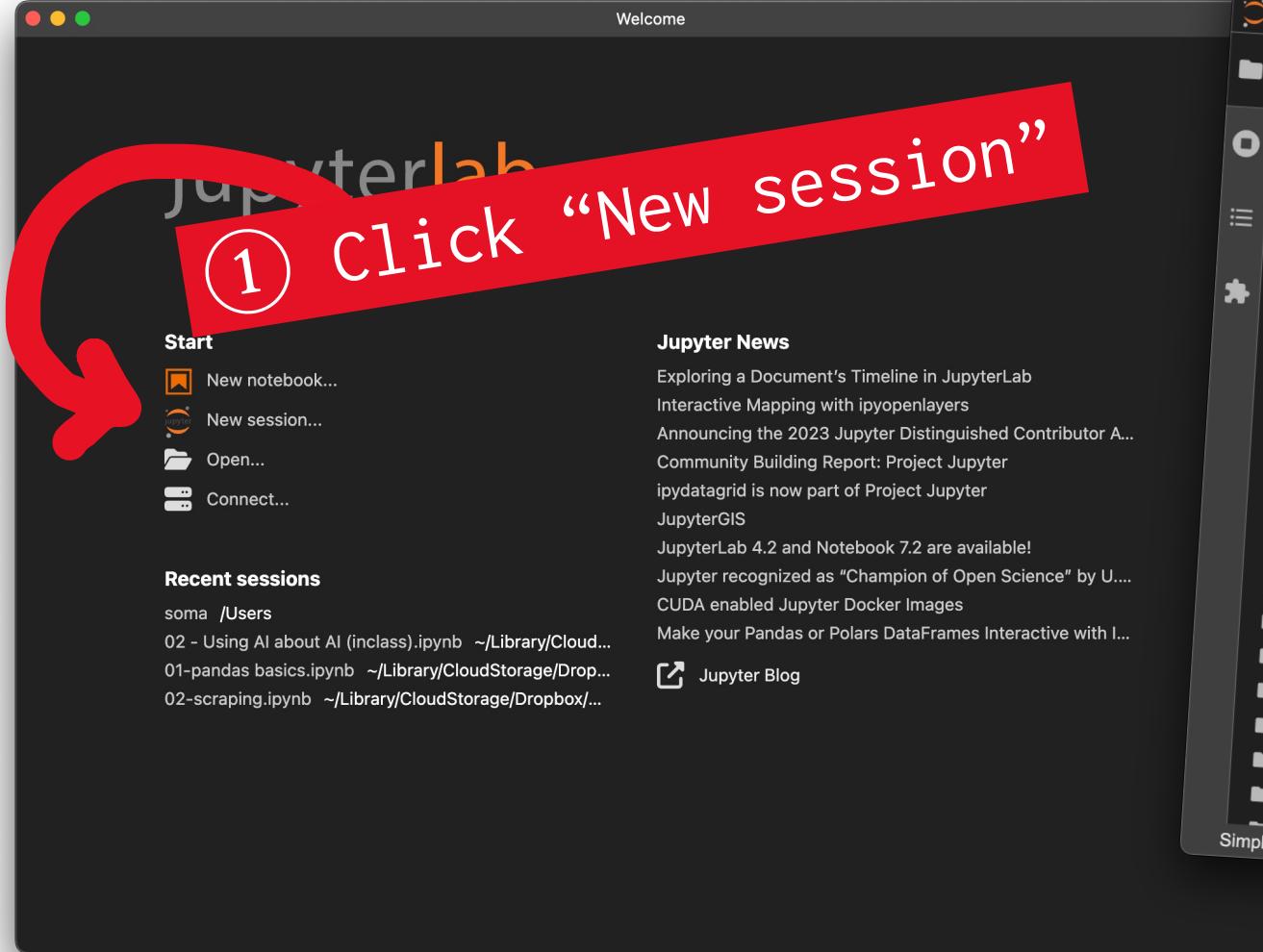
Don't run the
code, just think
and try to answer
the questions



Maybe easier on your
computer than phone?

<https://bit.ly/ds-dojo-python-quiz>

Create a new notebook



let's do some work!

time to try out some Python

Building Products

nytimes.com/interactive/2014/upshot/dialect-quiz-map.html

THE UPSHOT The New York Times GIVE THE TIMES Account

TheUpshot

How Y'all, Youse and You Guys Talk

By [Josh Katz](#) and [Wilson Andrews](#) Dec. 21, 2013

Share full article

What does the way you speak say about where you're from?
Answer all the questions below to see your personal dialect map.

QUESTION 1 OF 25

How would you address a group of two or more people?

- you all
- yous / youse
- you lot
- you guys
- you 'uns
- yinz
- you
- other
- y'all

Your Map

See the pattern of your dialect in the map below. Three of the most similar cities are shown.

Least similar Most similar

Show least similar SHARE YOUR MAP: [f](#) [t](#) [g](#)

Yonkers

Washington DC

Arlington

Can you build it?

Wikipedia日本語の方言

目次 非表示

ページ ノート 閲覧 編集 履歴表示 ツール

出典: フリー百科事典『ウィキペディア (Wikipedia)』

「本土方言」はこの項目へ転送されています。朝鮮語の本土方言については「朝鮮語#方言」をご覗ください。

この記事は検証可能な参考文献や出典が全く示されていないか、不十分です。出典を追加して記事の信頗性向上にご協力ください。(このテンプレートの使い方)

出典検索?: 「日本語の方言」 - ニュース・書籍・スカラー・CiNii・J-STAGE・NDL・dlib.jp・ジャパンサーチ・TWL (2021年2月)

日本語の方言 (にほんごのほうげん)、すなわち**日本語の地域変種** (地域方言)について記述する。

日本語は語彙・文法・音韻・アクセントなどあらゆる面で地方ごとの方言差が大きく、異なる地方に転居や旅行した際に、言葉が通じず苦労する場合が少なくない。日本語の方言は大きく「本土方言」と「琉球方言」に分かれ、それぞれがさらに細分化できる(区分章を参照)。明治以降、東京方言をもとに標準語の確立と普及が進められ、地方の方言はそれを阻害するものとして否定的にとらえられるようになった。共通語(第二次世界大戦後、標準語から共通語へと呼称が変わった)と、方言の共存が模索されるようになったが、実際には各地の伝統的な方言は急速に衰退・変質している(歴史・近代以降章を参照)。

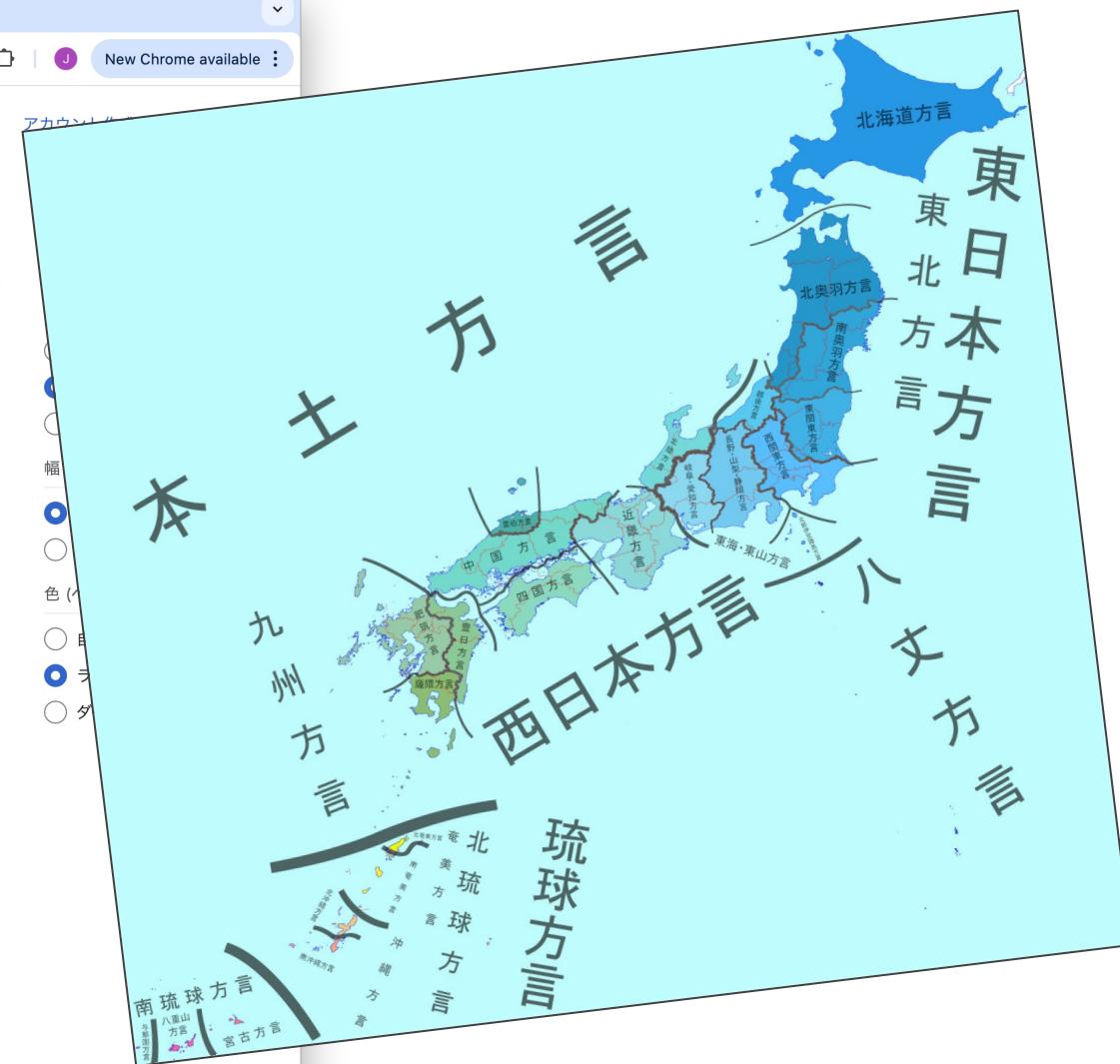
日本では「方言」という語は、「めんこい」「おもろい」「ばってん」のような共通語(標準語)とは異なる各地方独特の語彙や言い回し(「俚言」)、あるいはアクセントや発音の違い(いわゆる「なまり」)を指す場合が多い。しかし、言語学ではアクセント・音韻・文法などをすべてひっくるめ、「その地域社会の言語体系全体」を指すのが一般的である。すなわち、東京という一地域の日本語の体系ということで「東京方言」も当然存在する。

「言語」と「方言」

「琉球諸語#言語か方言か」も参照

本土方言と琉球方言は、文献時代に入る以前に分岐し、その後の往来も少なかったため、一聽する限り外国語同士に聞こえるほどの差が生じた。そのため、琉球方言を「琉球語」として、

だ	だ
じや	じや
や	や
いや	いや + や
や	や
だ + ジヤ	だ + ジヤ



claude.ai

The screenshot shows the Claude AI web interface. At the top, there's a pink header bar with the text "claude.ai". Below it is a browser window with the URL "claude.ai/new". The main content area is titled "Claude" and features a "Professional Plan" button. A large, friendly greeting "Good morning, j" is displayed, preceded by a small orange starburst icon. Below the greeting is a text input field asking "How can Claude help you today?". Underneath this, a section titled "Claude 3.5 Sonnet" is visible. A "Add content" button is present, along with three cards: "Provide stakeholder perspective", "Generate interview questions", and "Polish your prose". There's also a "Use a project" button and a "View all" link for recent chats. At the bottom left, there's a small circular icon with "JS" and a square icon.

claude.ai/new

Claude

Professional Plan

Good morning, j

How can Claude help you today?

Claude 3.5 Sonnet

Add content Use a project

Provide stakeholder perspective Generate interview questions Polish your prose

Your recent chats Show View all →

JS

Visit <https://bit.ly/ds-dojo-2024> for material

データ サイエンティスト DOJO

Jonathan Soma

js4571@columbia.edu

Knight Chair in Data Journalism, Columbia University