

How to data 101

How to scrape, clean, and visualise data.

Hi! We are Arran and Francesca.

The screenshot shows the homepage of the UCLAB website. The header features a teal background with a large, abstract, translucent wireframe mesh. On the left, there's a logo consisting of three overlapping triangles in green, blue, and purple. The navigation menu includes links for **ABOUT**, **PROJECTS**, **PUBLICATIONS**, and **PEOPLE**. A descriptive text block states: "UCLAB is a visualization research group situated between design, computing, and the humanities." The footer contains sections for **NEW COMPLEXITY** (with a sub-section for "Recent projects, publications and activities"), **THE FOLD** (with a diagram showing various line patterns), a map of Germany with city labels, and a book titled "Visuals for Language". The footer also includes the FH;P logo.

URBAN
COMPLEXITY
LAB

ABOUT PROJECTS PUBLICATIONS PEOPLE

UCLAB is a visualization research group situated between design, computing, and the humanities.

FH;P

NEW COMPLEXITY
Recent projects, publications and activities

THE FOLD

Visuals for Language

Hamburg, Berlin, Dresden, Leipzig, Königsberg, Danzig, Breslau, Göttingen, Frankfurt am Main, Nürnberg, Freiburg, Stuttgart, München

VISUALS FOR

Nice to meet you, what about you?

Scrape. theory + python

Clean. python

Visualise. theory + rawGraphs

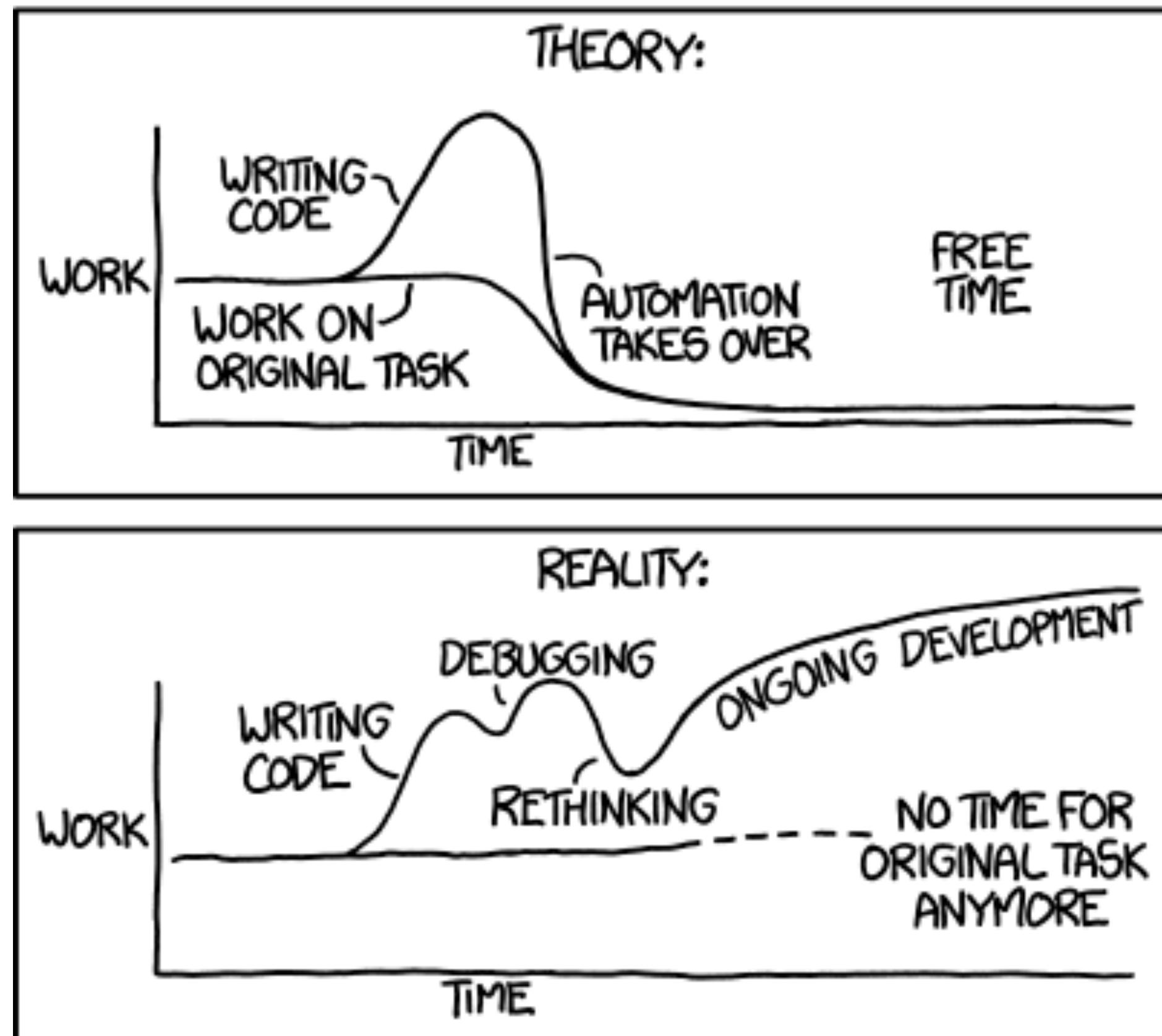
How to scrape

Way faster (once it works)

Web scraping means accessing
and extracting data from
websites. it can be done
manually (by opening links,
copying and pasting information
and so on) or automatically via
crawlers and bots.

Let's create our own data!

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



It's a trap!

<https://xkcd.com/>

SOME FUNCTIONALITY ON THIS SITE REQUIRES YOUR CONSENT FOR COOKIES TO WORK PROPERLY.

I consent to cookies I want more information

Rotten Tomatoes®

Search movies, TV, actors, more... 

MOVIES & DVDS ▾ TV ▾ NEWS ▾ TICKETS & SHOWTIMES

What's the Tomatometer®? Critics SIGN UP | LOG IN

TRENDING ON RT Best Amazon Shows Podcast: Avatar 2021 TV Premiere Dates Best Netflix Movies 

IN THEATERS

- Opening This Week
- Top Box Office
- Coming Soon
- Certified Fresh Movies

DVD & STREAMING MOVIES

- Browse All 
- Top DVD & Streaming
- New Releases
- Coming Soon
- Certified Fresh Movies

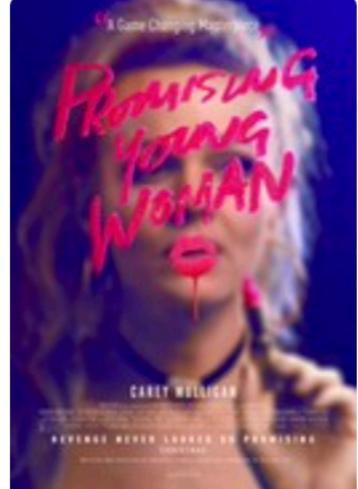
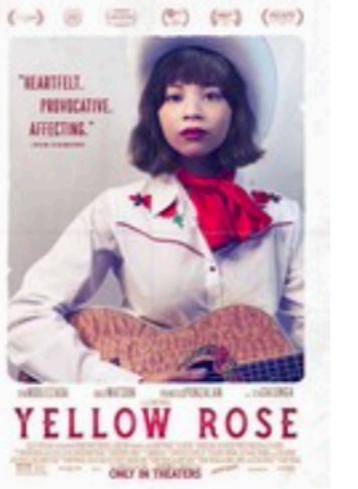
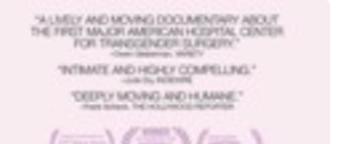
TV

- New TV Tonight
- Most Popular TV on RT
- Certified Fresh TV - 2020

BROWSE ALL

PROVIDERS ▾  0% —  100% ▾ Genre: All ▾ Sort By: Release Date ▾

Showing 32 of 22711  

 The Emoji Story (Picture Character)  91% Available Dec 22	 Promising Young Woman  91% Available Jan 15	 Yellow Rose  86% Available Jan 5	 American Skin  31% Available Jan 15
 Streaming Exclusively December 25 Disney+ Soul	 "A MESSY AND MOVING DOCUMENTARY ABOUT THE FIRST MAJOR AMERICAN HOSPITAL CENTER FOR TRANSGENDER PATIENTS." "INTIMATE AND HONKY COMPELLING." "DEEPLY MOVING AND HUMANE." American Skin	 Yellow Rose  86% Available Jan 5	 American Skin  31% Available Jan 15

SOME FUNCTIONALITY ON THIS SITE REQUIRES YOUR CONSENT FOR COOKIES TO WORK PROPERLY.

I consent to cookies I want more information

PROVIDERS 0% — 100%

What's the Tomatometer®? Critics SIGN UP | LOG IN

Search movies, TV, actors, more... MOVIES & DVDS ▾ TV ▾ NEWS ▾ TICKETS & SHOWTIMES

Showing 32 of 22711

Coming Soon Certified Fresh Movies

BROWSE ALL

PROVIDERS 0% — 100% ▾ Genre: All ▾ Sort By: Release Date ▾

Showing 32 of 22711

DVD & STREAMING MOVIES

Browse All Top DVD & Streaming New Releases Coming Soon Certified Fresh Movies

TV

New TV Tonight Most Popular TV on RT Certified Fresh TV - 2020

The Emoj Story (Picture Character) 91% Available Dec 22

Promising Young Woman 91% Available Jan 15

Yellow Rose 86% Available Jan 5

American Skin 31% Available Jan 15

CERTIFIED FRESH

Streaming Exclusively December 25 Disney+ COOL

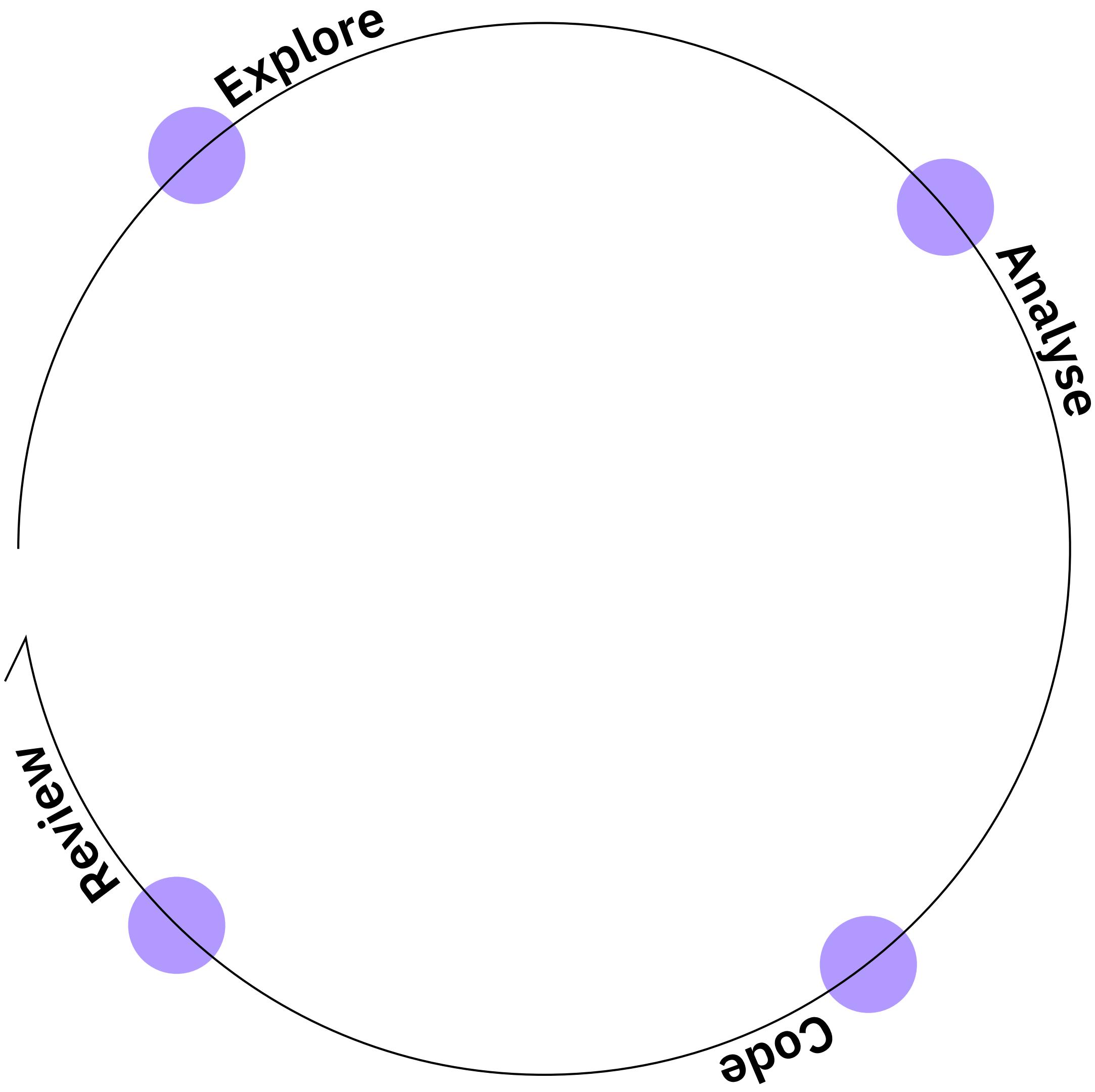
Professio... Scouting... W...
"A MELT AND MOVING DOCUMENTARY ABOUT THE FIRST MAJOR AMERICAN HOSPITAL CENTER FOR TRANSSEXUAL SURGERY. INTIMATE AND HONKY COMPELLING." DEEPLY MOVING AND HUMANE."

URBAN COMPLEXITY LAB

super important!

Necessary ingredients:

- a good amount of time
- interesting website source
- a solid scraping strategy
- python Selenium
- chrome web driver
- python pandas



10 minutes break :)

Hands on tutorials.

<https://github.com/uclab-potsdam/filmuni-workshop/>

15 minutes break :)

Brainstorming and groups forming

<https://docs.google.com/spreadsheets/film-uni-workshop>

Ideas brainstorming (10 minutes)

Ideas voting (10 minutes)

Groups forming (10 minutes)