

An illustration of a woman with dark hair, wearing a light blue headscarf and a white long-sleeved top. She is looking down at a small, brown, textured object she is holding with both hands. The background is a dark, textured brown with some geometric shapes. The overall style is a soft, painterly illustration.

Practical code archaeology

Judith van Stegeren
@jd7h

PyData Eindhoven 2022



Tweet



Vicki @vboykis · Jul 27



Gonna start a conference called [#NormIPS](#) that's just presentations of middlebrow ML topics. "how to structure Python packages 2022", "how many k-folds is too many", "how to make the browser pop-up come up when the notebook is done running", "putting features in Postgres", etc.



72



156



1,195



Adam Lauretig II @lauretig@mastodon.social



@lauretig

Replying to [@vboykis](#)

This is great.

My contribution: "Figuring out what that notebook from two years ago did: practical code archeology"

2:13 AM · Jul 27, 2022 · Twitter Web App

2 Retweets

1 Quote Tweet

54 Likes



CODE ARCHAEOLOGY

CODE ARCHAEOLOGY EVERYWHERE

A photograph showing two individuals, likely students or researchers, working on an excavation of a human skeleton in a dirt pit. One person, with bright green hair, is using a small red-handled tool to carefully remove soil from the ribcage area. The other person, with brown hair in a ponytail, is also working on the skeleton. The skull is visible on the left, and the rest of the skeleton is laid out on the ground. The scene is outdoors, with sunlight casting shadows.

**THIS YEAR'S
DATA SCIENTISTS**

**SCRIPT FROM
LAST YEAR'S DATA
SCIENTISTS**

The background of the image is a photograph of the Terracotta Army, showing rows of ancient Chinese warrior statues in a museum setting. The statues are made of terracotta and are arranged in long, winding lines, creating a sense of depth and scale. The lighting is warm, highlighting the texture of the clay and the details of the armor and weapons.

OLD REPO

FIRST PROTOTYPE
BY INTERN

TEMP.IPYNB

TRAINING.IPYNB

ETL.IPYNB

ANALYSIS.IPYNB

ANALYSIS.IPYNB

ANALYSIS1.IPYNB

DASHBOARD.IPYNB

DASHBOARDING.IPYNB

DEEPNOTE

AWS SAGEMAKER



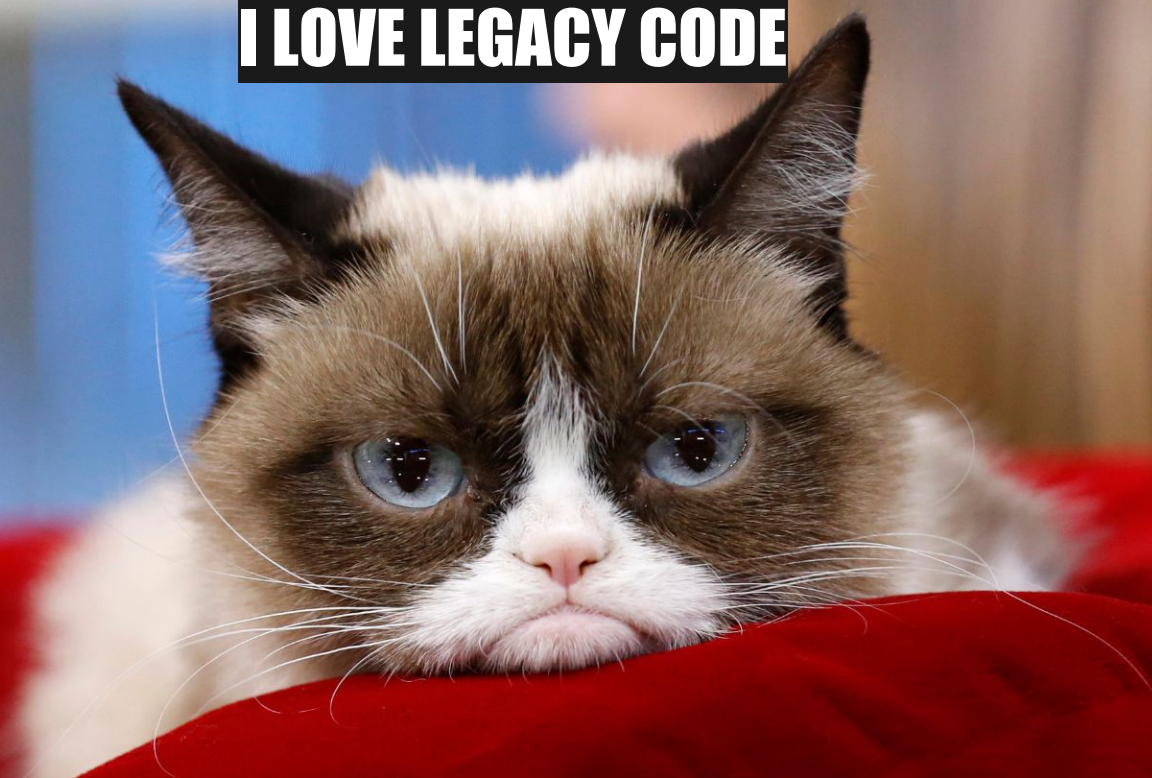
**Good Mood: I excavate
and study important
legacy production code**

**BAD MOOD: I DIG UP OTHER
PEOPLE'S TRASH AND ORGANIZE IT**

imgflip.com



I LOVE LEGACY CODE



IT MAKES PEOPLE CRY

1. Talking beats debugging



2. Always start with
the simplest option



3. Invest in your detective skills



The more programming I do, the more thing I run into where

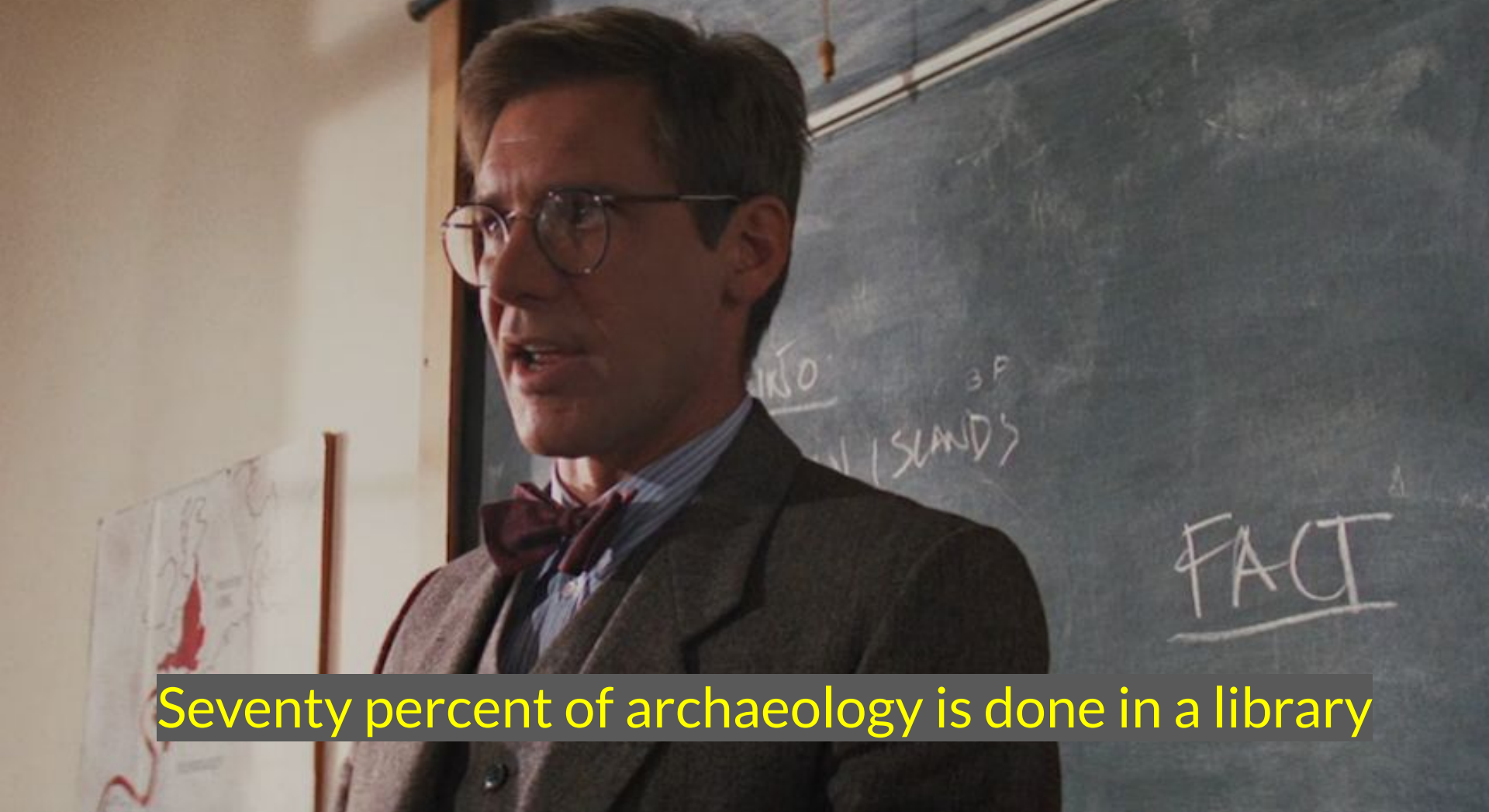
- I don't know
- Google doesn't know
- my colleagues don't know
- we gotta do it anyway

When this happens, I think "Right, this is why they pay a human with a brain who can investigate and learn"



Look up
"How to be a wizard programmer"
by Julia Evans





Seventy percent of archaeology is done in a library

```
product_names = ["A", "C", "D"]
```

What happened to B?

Most important object class
at a car dealership:

Class Bike

Table names in data warehouse/lake:

Licenses

LicenseInfo

LicenseData

LicensePlates

LicenseRegistrations

git blame
git blame
git blame

```
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 1) xoreos-tools README
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 2) =====
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 3)
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 4) A collection of tools to help with the reverse-engineering of BioWare's
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 5) Aurora engine games. xoreos-tools is part of the xoreos project; please
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 6) see the [xoreos website](https://xoreos.org/) and its [GitHub
7923656e (Sven Hesse 2015-07-14 02:04:29 +0200 7) repositories](https://github.com/xoreos) for details, especially [the
7923656e (Sven Hesse 2015-07-14 02:04:29 +0200 8) main README.md](https://github.com/xoreos/xoreos/blob/master/README.md).
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 9)
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 10) The tools included here are licensed under the terms of the [GNU General
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 11) Public License version 3](https://www.gnu.org/licenses/agpl-3.0.html)
4bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 12) or (at your option) any later version.
0b52ee55 (Sven Hesse 2014-10-20 18:13:26 +0200 13)
0b52ee55 (Sven Hesse 2014-10-20 18:13:26 +0200 14) Currently, the following tools are included:
0b52ee55 (Sven Hesse 2014-10-20 18:13:26 +0200 15)
0b52ee55 (Sven Hesse 2014-10-20 18:13:26 +0200 16) * gff2xml: Convert BioWare GFF to XML
ecdd8d48 (Sven Hesse 2015-05-01 14:15:27 +0200 17) * tlk2xml: Convert BioWare TLK to XML
38d9b032 (Sven Hesse 2016-04-14 16:23:01 +0200 18) * ssf2xml: Convert BioWare SSF to XML
5ad3c6dd (patrick 2019-09-28 21:07:42 +0200 19) * fev2xml: Convert FMOD FEV to XML
b358a873 (patrick 2018-09-28 18:30:20 +0200 20) * xml2gff: Convert XML back to BioWare GFF
3332d1ec (Sven Hesse 2015-09-15 15:50:48 +0200 21) * xml2tlk: Convert XML back to BioWare TLK
b2aca536 (Sven Hesse 2016-04-19 19:58:26 +0200 22) * xml2ssf: Convert XML back to BioWare SSF
b5d911d2 (Sven Hesse 2015-05-29 18:28:06 +0200 23) * convert2da: Convert BioWare 2DA/GDA to 2DA/CSV
aba905ae (Sven Hesse 2015-11-14 21:50:23 +0100 24) * fixpremiumgff: Repair BioWare GFF files in NWN premium module HAKs
34349021 (Sven Hesse 2018-10-28 11:34:02 +0100 25) * fixnwn2xml: Convert Obsidian NWN2 XML to valid XML
f0262c75 (Sven Hesse 2014-10-20 18:16:27 +0200 26) * unerf: Extract BioWare ERF archives
ecdd8d48 (Sven Hesse 2015-05-01 14:15:27 +0200 27) * unherf: Extract BioWare HERF archives
f0447410 (Sven Hesse 2014-10-20 18:17:12 +0200 28) * unrिम: Extract BioWare RIM archives
5521767e (Sven Hesse 2014-10-20 18:17:48 +0200 29) * unnds: Extract Nintendo DS roms
b6841bf3 (Sven Hesse 2015-05-05 20:57:46 +0200 30) * unnsbtx: Extract Nintendo NSBTX textures into TGA images
5a22c459 (Sven Hesse 2014-10-20 18:18:39 +0200 31) * unkeybif: Extract BioWare KEY/BIF archives
4c02921c (Sven Hesse 2018-07-21 15:47:32 +0200 32) * unobb: Extract Aspyr's OBB virtual filesystem
bc45555a (patrick 2018-09-27 01:19:02 +0200 33) * untws: Extract CDProjectRed's TheWitcherSave archives
42271031 (patrick 2018-06-25 16:25:52 +0200 34) * erf: Create BioWare ERF archives
9ca0557c (patrick 2019-04-20 16:56:05 +0200 35) * rim: Create BioWare RIM archives
bc45555a (patrick 2018-09-27 01:19:02 +0200 36) * tws: Create CDProjectRed TheWitcherSave archives
0d4d3623 (patrick 2018-10-14 17:35:08 +0200 37) * keybif: Create BioWare KEY/BIF archives
9f15fad8 (Sven Hesse 2015-05-02 21:41:05 +0200 38) * desmall: Decompress "small" (Nintendo DS LZSS types 0x00 and 0x10) files
```

tig




```
$ cd code/bookwyrms/
$ tig
```

```
2022-11-01 20:53 -0700 Mouse Reeve      M  [main] {origin/main} {origin/HEAD} Merge pull
2022-11-01 20:43 -0700 Mouse Reeve      |  o Python formatting
2022-11-01 19:26 -0700 Mouse Reeve      |  o Mark imports as complete
2022-11-01 18:59 -0700 Mouse Reeve      |  o Adds import admin view
2022-11-01 18:45 -0700 Mouse Reeve      |  o Use elided page range on user admin view
2022-11-01 18:43 -0700 Mouse Reeve      M  Merge pull request #2326 from bookwyrms-socia
2022-11-01 18:20 -0700 Mouse Reeve      |  o Updates test mock
2022-11-01 18:17 -0700 Mouse Reeve      |  o Fixes rate limiter
2022-11-01 18:08 -0700 Mouse Reeve      |  o Send confirmation emails directly, rather th
2022-11-01 18:18 -0700 Mouse Reeve      M  Merge pull request #2325 from bookwyrms-soc
2022-11-01 18:04 -0700 Mouse Reeve      |  o Only show editions with the author on the
2022-11-01 17:15 -0700 Mouse Reeve      M  Merge pull request #2322 from obrenckle/me
2022-10-27 14:04 +0200 Olivier Brencklé |  o :children_crossing: Close dropdown when openin
2022-10-24 19:34 +0200 Olivier Brencklé |  o :lipstick: Set dropdown menu below modals on m
2022-10-21 14:14 -0700 Mouse Reeve      M  Merge pull request #2321 from bookwyrms-social/
2022-10-20 20:57 -0700 Mouse Reeve      M  Merge branch 'main' into dependabot/pip/djan
2022-10-21 03:40 +0000 dependabot[bot]  |  o Bump django from 3.2.15 to 3.2.16
2022-10-21 14:13 -0700 Mouse Reeve      |  o Adds 2fa locales
2022-10-20 20:56 -0700 Mouse Reeve      M  Merge pull request #2320 from bookwyrms-socia
2022-10-20 20:44 -0700 Mouse Reeve      |  o Disable wayward stylelint error
2022-10-20 07:45 -0700 Mouse Reeve      |  o Remove extra whitespace in announcements
2022-10-20 20:40 -0700 Mouse Reeve      M  Merge pull request #2294 from hughrun/otp
2022-10-15 07:35 +1100 Hugh Rundle      |  o handle 2fa user exception properly
2022-10-15 07:23 +1100 Hugh Rundle      |  o pylint is being pedantic
2022-10-15 07:11 +1100 Hugh Rundle      |  o return Bad Request if 2fa user does not exis
[main] 6eab1fdd6d97fa05e0d307627f06aa3d0ebb3b5a - commit 4 of 7611 0%
```

```
$ cd code/bookwyrm/
$ tig
```

```
2022-11-01 20:53 -0700 Mouse Reeve
2022-11-01 20:43 -0700 Mouse Reeve
2022-11-01 19:26 -0700 Mouse Reeve
2022-11-01 18:59 -0700 Mouse Reeve
2022-11-01 18:45 -0700 Mouse Reeve
2022-11-01 18:43 -0700 Mouse Reeve
2022-11-01 18:20 -0700 Mouse Reeve
2022-11-01 18:17 -0700 Mouse Reeve
2022-11-01 18:08 -0700 Mouse Reeve
2022-11-01 18:18 -0700 Mouse Reeve
2022-11-01 18:04 -0700 Mouse Reeve
2022-11-01 17:15 -0700 Mouse Reeve
2022-10-27 14:04 +0200 Olivier Brencklé
2022-10-24 19:34 +0200 Olivier Brencklé
2022-10-21 14:14 -0700 Mouse Reeve
2022-10-20 20:57 -0700 Mouse Reeve
2022-10-21 03:40 +0000 dependabot[bot]
2022-10-21 14:13 -0700 Mouse Reeve
2022-10-20 20:56 -0700 Mouse Reeve
2022-10-20 20:44 -0700 Mouse Reeve
2022-10-20 07:45 -0700 Mouse Reeve
2022-10-20 20:40 -0700 Mouse Reeve
2022-10-15 07:35 +1100 Hugh Rundle
2022-10-15 07:23 +1100 Hugh Rundle
2022-10-15 07:11 +1100 Hugh Rundle
```

```
[main] 6eab1fdd6d97fa05e0d307627f06aa3d0ebb3b5a - c0%
```

```
commit 6eab1fdd6d97fa05e0d307627f06aa3d0ebb3b5a
```

```
Author: Mouse Reeve <mouserreeve@riseup.net>
```

```
AuthorDate: Tue Nov 1 18:59:04 2022 -0700
```

```
Commit: Mouse Reeve <mouserreeve@riseup.net>
```

```
CommitDate: Tue Nov 1 19:05:51 2022 -0700
```

```
    Adds import admin view
```

```
---
```

```
bookwyrm/models/import_job.py      | 9 ++
.../settings/imports.html          | 71 ++++++++
.../settings/layout.html           | 6 +
bookwyrm/urls.py                   | 3 +
bookwyrm/views/__init__.py         | 1 +
bookwyrm/views/admin/imports.py    | 33 +++++
6 files changed, 123 insertions(+)
```

```
diff --git a/bookwyrm/models/import_job.py b/bookwyr
index 556f133f9..7547cfcd2 100644
```

```
--- a/bookwyrm/models/import_job.py
```

```
+++ b/bookwyrm/models/import_job.py
```

```
@@ -48,6 +48,15 @@ class ImportJob(models.Model):
     """items that haven't been processed yet"""
     return self.items.filter(fail_reason__isnu
```

```
+     @property
```

```
[diff] 6eab1fdd6d97fa05e0d307627f06aa3d0ebb3b5a 12%
```

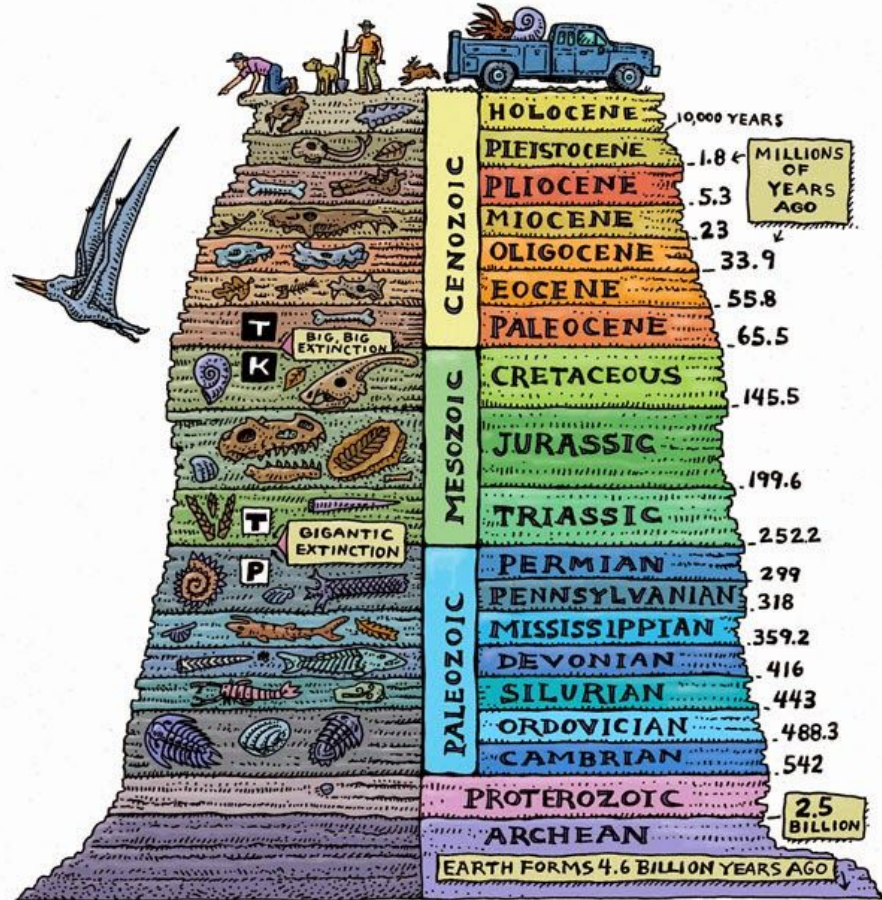
```
$ cd code/bookwyrms/
$ tig README.md
```

```
2022-07-11 15:55 -0700 Mouse Reeve      o Update README.md
2022-07-11 14:47 -0700 Mouse Reeve      o Update README.md
2022-05-15 13:01 +0200 Ricardo Rodríguez o docs: Fix broken links in README
2022-04-08 13:59 -0700 Mouse Reeve      o Consistent formatting for "BookWyrms" name
2022-02-28 21:23 +0100 Strubbl         o Update README.md
2021-04-13 12:17 -0700 Mouse Reeve      o wording fix
2021-04-13 12:16 -0700 Mouse Reeve      o more links
2021-04-13 12:15 -0700 Mouse Reeve      o Fixes links from readme to docs
2021-04-13 10:26 -0700 Mouse Reeve      o Updates README for new docs website
2021-04-12 15:11 -0700 Mouse Reeve      o Fixes markdown formatting
2021-04-12 20:44 +0200 tofuwabohu       o Indentation in md
2021-04-12 20:12 +0200 tofuwabohu       M- Resolve merge conflicts
2021-04-12 09:34 -0700 Mouse Reeve      | M- Merge pull request #871 from SavinaRojas/ease-production
2021-03-30 09:28 -0400 Pablo Barton     | | o Easier to deploy by avoiding merge conflicts in production
2021-04-12 11:56 -0400 n2nol           | | o update links in README and {bookwyrms,celerywyrms}/settings.
2021-04-11 19:03 +0200 Your Name        | o Adds final newlines
2021-04-11 18:58 +0200 Your Name        | o Fixes link
2021-04-11 18:57 +0200 Your Name        | o Puts Installation and Contributing instruction in their own
2021-04-10 12:58 +0200 Fabien Basmaison M- Merge branch 'main' into frontend
2021-04-09 17:44 -0400 n2nol           | o remove duplicated install instructions in README.md
2021-04-06 11:54 +0200 Fabien Basmaison | o [assets] Allow to run `./bw-dev collectstatic` every time a
2021-03-23 13:38 -0700 Mouse Reeve      o Updates readme
2021-03-21 14:31 -0700 Mouse Reeve      o Copies over the production README
2021-03-21 09:20 -0700 Mouse Reeve      o Updates readme
2021-03-20 15:30 +0100 Fabien Basmaison o Fix indentation in README.md.
```

```
[main] e3dbe5a1420b271462b20d25a88dba0b638f773d - commit 1 of 76
```

```
32%
```


Stratigraphy



```
$ git blame --color-by-age
```

```
$ git config --local --add color.blame.highlightRecent  
"red,6 year ago,yellow,5 year ago,green,4 year  
ago,blue,3 year ago,cyan,2 year ago,magenta,1 year  
ago,white"
```

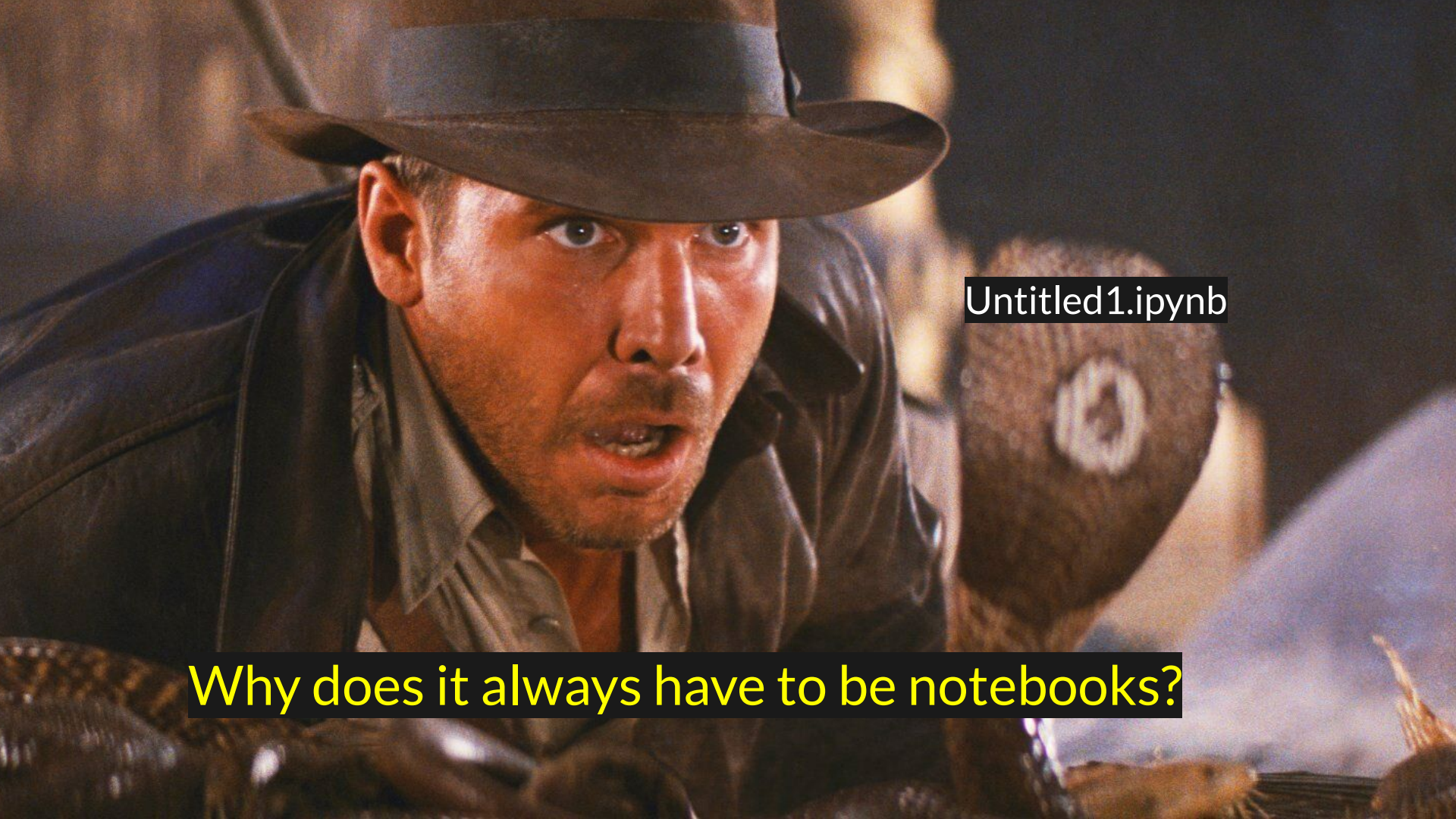
```
$ git config --local --add color.blame.highlightRecent  
"red,6 year ago,yellow,5 year ago,green,4 year  
ago,blue,3 year ago,cyan,2 year ago,magenta,1 year  
ago,white"
```

```
$ git config --local --add color.blame.highlightRecent  
"red,6 year ago,yellow,5 year ago,green,4 year  
ago,blue,3 year ago,cyan,2 year ago,magenta,1 year  
ago,white"
```

```
$ git config --local --add color.blame.highlightRecent  
"red,6 year ago,yellow,5 year ago,green,4 year  
ago,blue,3 year ago,cyan,2 year ago,magenta,1 year  
ago,white"
```

```
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 1) xoreos-tools README
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 2) =====
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 3)
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 4) A collection of tools to help with the reverse-engineering of BioWare's
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 5) Aurora engine games. xoreos-tools is part of the xoreos project; please
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 6) see the [xoreos website](https://xoreos.org/) and its [GitHub
923656e (Sven Hesse 2015-07-14 02:04:29 +0200 7) repositories](https://github.com/xoreos) for details, especially [the
923656e (Sven Hesse 2015-07-14 02:04:29 +0200 8) main README.md](https://github.com/xoreos/xoreos/blob/master/README.md).
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 9)
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 10) The tools included here are licensed under the terms of the [GNU General
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 11) Public License version 3](https://www.gnu.org/licenses/agpl-3.0.html)
bd5b960 (Sven Hesse 2014-10-20 17:23:32 +0200 12) or (at your option) any later version.
b52ee55 (Sven Hesse 2014-10-20 18:13:26 +0200 13)
b52ee55 (Sven Hesse 2014-10-20 18:13:26 +0200 14) Currently, the following tools are included:
b52ee55 (Sven Hesse 2014-10-20 18:13:26 +0200 15)
b52ee55 (Sven Hesse 2014-10-20 18:13:26 +0200 16) * gff2xml: Convert BioWare GFF to XML
cdd8d48 (Sven Hesse 2015-05-01 14:15:27 +0200 17) * tlk2xml: Convert BioWare TLK to XML
8d9b032 (Sven Hesse 2016-04-14 16:23:01 +0200 18) * ssf2xml: Convert BioWare SSF to XML
ad3c6dd (patrick 2019-09-28 21:07:42 +0200 19) * fev2xml: Convert FMOD FEV to XML
358a873 (patrick 2018-09-28 18:30:20 +0200 20) * xml2gff: Convert XML back to BioWare GFF
332d1ec (Sven Hesse 2015-09-15 15:50:48 +0200 21) * xml2tlk: Convert XML back to BioWare TLK
2aca536 (Sven Hesse 2016-04-19 19:58:26 +0200 22) * xml2ssf: Convert XML back to BioWare SSF
5d911d2 (Sven Hesse 2015-05-29 18:28:06 +0200 23) * convert2da: Convert BioWare 2DA/GDA to 2DA/CSV
ba905ae (Sven Hesse 2015-11-14 21:50:23 +0100 24) * fixpremiumgff: Repair BioWare GFF files in NWN premium module HAKs
4349021 (Sven Hesse 2018-10-28 11:34:02 +0100 25) * fixnwn2xml: Convert Obsidian NWN2 XML to valid XML
0262c75 (Sven Hesse 2014-10-20 18:16:27 +0200 26) * unerf: Extract BioWare ERF archives
cdd8d48 (Sven Hesse 2015-05-01 14:15:27 +0200 27) * unherf: Extract BioWare HERF archives
0447410 (Sven Hesse 2014-10-20 18:17:12 +0200 28) * unrim: Extract BioWare RIM archives
521767e (Sven Hesse 2014-10-20 18:17:48 +0200 29) * unnnds: Extract Nintendo DS roms
6841bf3 (Sven Hesse 2015-05-05 20:57:46 +0200 30) * unnstbx: Extract Nintendo NSBTX textures into TGA images
a22c459 (Sven Hesse 2014-10-20 18:18:39 +0200 31) * unkeybif: Extract BioWare KEY/BIF archives
c02921c (Sven Hesse 2018-07-21 15:47:32 +0200 32) * unobb: Extract Aspyr's OBB virtual filesystem
c45555a (patrick 2018-09-27 01:19:02 +0200 33) * untws: Extract CDProjectRed's TheWitcherSave archives
2271031 (patrick 2018-06-25 16:25:52 +0200 34) * erf: Create BioWare ERF archives
ca0557c (patrick 2019-04-20 16:56:05 +0200 35) * rim: Create BioWare RIM archives
c45555a (patrick 2018-09-27 01:19:02 +0200 36) * tws: Create CDProjectRed TheWitcherSave archives
d4d3623 (patrick 2018-10-14 17:35:08 +0200 37) * keybif: Create BioWare KEY/BIF archives
f15fad8 (Sven Hesse 2015-05-02 21:41:05 +0200 38) * desmall: Decompress "small" (Nintendo DS LZSS, types 0x00 and 0x10) files
f467344 (Sven Hesse 2015-07-22 18:24:22 +0200 39) * xoreostext2tga: Convert BioWare's texture formats into TGA
e707916 (Sven Hesse 2015-05-09 00:17:20 +0200 40) * nbfs2tga: Convert Nintendo's raw NBFS images into TGA
bf84c48 (Sven Hesse 2015-05-16 19:54:25 +0200 41) * ncgr2tga: Convert Nintendo's NCGR images into TGA
920b91e (Sven Hesse 2015-05-25 15:48:04 +0200 42) * cbgt2tga: Convert CBGT images into TGA
530ae29 (Sven Hesse 2015-05-25 18:59:05 +0200 43) * cdpth2tga: Convert CDPth depth images into TGA
6876fd4 (Sven Hesse 2015-10-03 20:00:42 +0200 44) * ncstdis: Disassemble NWScript bytecode
3122d55 (patrick 2019-03-24 18:22:34 +0100 45) * ncstdcomp: Decompile NWScript bytecode
```

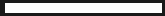
git log -p
git bisect
find
grep
netstat
tcpdump



Untitled1.ipynb

Why does it always have to be notebooks?

“Just run it”



“Just run it”





“Just read it”

An illustration of a woman with dark hair, wearing a blue headscarf and a white robe, looking down at a small, glowing robot. The robot has a white, wireframe-like head and a body with orange and blue segments. They are in a garden with green plants and a stone building in the background. A speech bubble from the robot contains the text: "This could be an old prototype... OR MAYBE I'M HOLDING IT UPSIDE DOWN, I DON'T KNOW."

This could be an old prototype...
OR MAYBE I'M HOLDING IT UPSIDE
DOWN, I DON'T KNOW.

**Try to answer the
basic questions first**

**Who
What
When
Why**

Reading strategies

- I/O
- dependencies/imports
- code structure
- class names
- logging
- docstrings
- comments
- use your IDE





Do a pairing session

When all else fails





?





I CAN JUST MAKE OUT SOME WORN
MARKINGS...

Thank you :)

Judith van Stegeren

@jd7h

judith@jd7h.com

www.judithvanstegeren.com