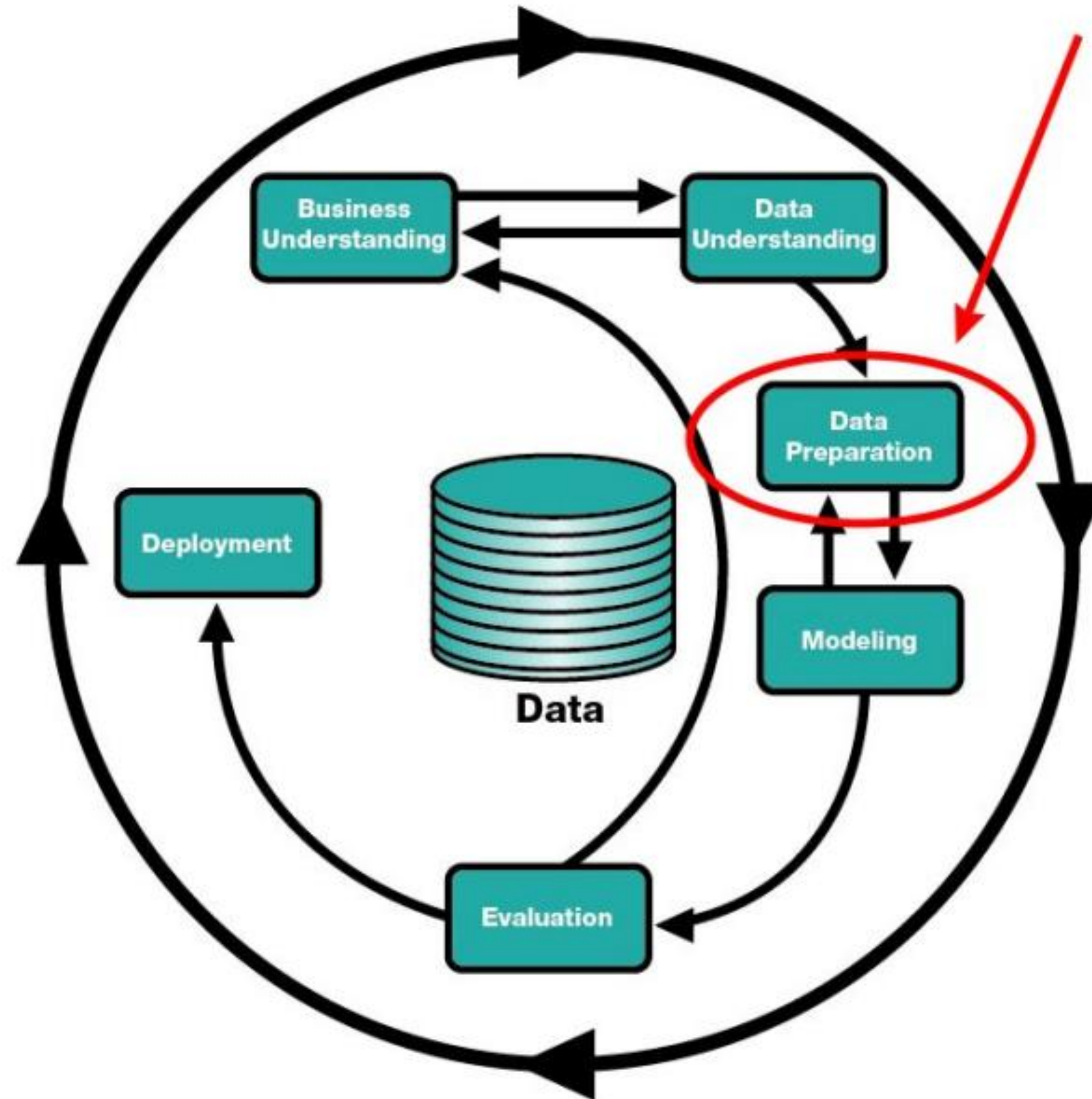# Data Preparation with
Python
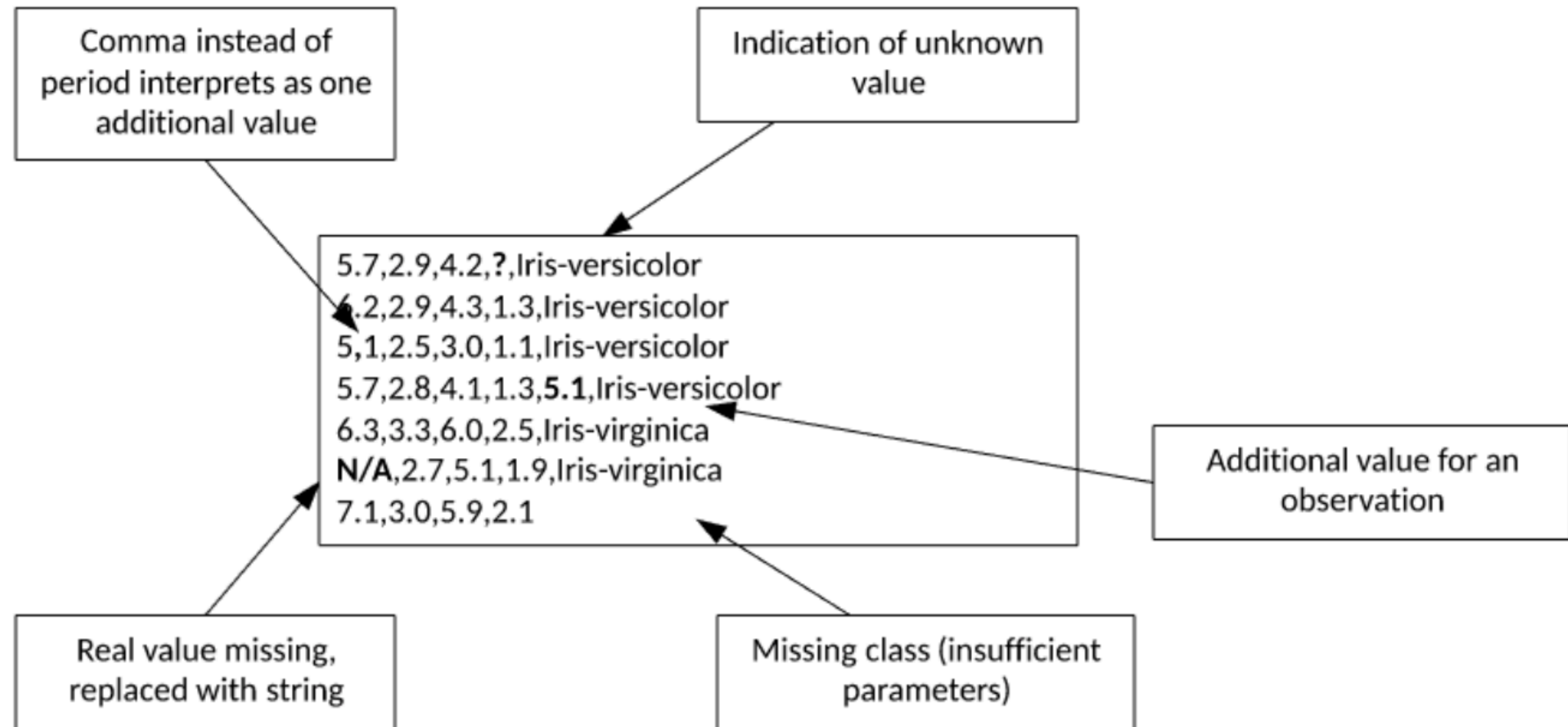
Galuh Sahid

Data Engineer at Gojek | Mozilla Tech Speaker

@galuhsahid | github.com/galuhsahid | speakerdeck.com/galuhsahid

@galuhsahid

	sepal length	sepal width	petal length	petal width	target
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

Comma instead of period interprets as one additional value

Indication of unknown value

```
5.7,2.9,4.2,?,Iris-versicolor
4.2,2.9,4.3,1.3,Iris-versicolor
5,1,2.5,3.0,1.1,Iris-versicolor
5.7,2.8,4.1,1.3,5.1,Iris-versicolor
6.3,3.3,6.0,2.5,Iris-virginica
N/A,2.7,5.1,1.9,Iris-virginica
7.1,3.0,5.9,2.1
```

Additional value for an observation

Real value missing, replaced with string

Missing class (insufficient parameters)

```json
        "to_user_id_str": "101622611",
        "source": "&lt;a href=&quot;https://chrome.google.com/extensions/detail/encaiiljifbdbjlphpgpiimidegddhic&quot;
        rel=&quot;nofollow&quot;&gt;Silver Bird&lt;/a&gt;"
    },
    {

        "from_user_id_str": "17777784",
        "profile_image_url": "http://a1.twimg.com/profile_images/1323147222/IMG0316A_-_Copy_normal.jpg",
        "created_at": "Thu, 30 Jun 2011 06:41:16 +0000",
        "from_user": "vikaskanani",
        "id_str": "86323395834290176",
        "metadata": {
            "result_type": "recent"
        },
        "to_user_id": 101622611,
        "text": "@pareshmayani ya right we are too much behind, google street view is not here, what else?",
        "id": 86323395834290180,
        "from_user_id": 17777784,
        "to_user": "pareshmayani",
        "geo": null,
        "iso_language_code": "en",
        "to_user_id_str": "101622611",
        "source": "&lt;a href=&quot;https://chrome.google.com/extensions/detail/encaiiljifbdbjlphpgpiimidegddhic&quot;
        rel=&quot;nofollow&quot;&gt;Silver Bird&lt;/a&gt;"
    }
],
"max_id": 87807860502831100,
"since_id": 0,
"refresh_url": "?since_id=87807860502831104&q=pareshmayani",
"results_per_page": 15,
"page": 1,
"completed_in": 0.186334,
"since_id_str": "0",
"max_id_str": "87807860502831104",
"query": "pareshmayani"
}
```

2016 (Source A)

Country	GDP (USD)
Indonesia	22429285
India	23481299

2017 (Source B)

Country	GDP (USD Million)
Indonesia	24
India	28

pandas

panel data

First ever black hole image released

By Pallab Ghosh
Science correspondent, BBC News

🕑 10 April 2019 　　f 💬 🐦 ✉ ⤴ Share

EHT

The first ever picture of a black hole: It's surrounded by a halo of bright gas

Astronomers have taken the first ever image of a black hole, which is located in a distant galaxy.

It measures 40 billion km across - three million times the size of the Earth - and has been described by scientists as "a monster".

The black hole is 500 million trillion km away and was photographed by a network of eight telescopes across the world.

Python Used to Take Photo of Black Hole

⏱ April 11, 2019   🗀 Python   🏷 Python   👤 Mike

Scientists have used a new algorithm to take a photo of a black hole. One of the most exciting parts about it to me is that they used a lot of Python libraries to do the magic.

Here's a list mentioned in their paper:

- Numpy (van der Walt et al. 2011)
- Scipy (Jones et al. 2001)
- Pandas (McKinney 2010)
- Jupyter (Kluyver et al. 2016)
- Matplotlib (Hunter 2007).
- Astropy (The Astropy Collaboration et al. 2013, 2018)

They also used their own custom Python code which is available on Github

https://github.com/galuhsahid/data-preparation-with-python