

# RubyData エコシステムへの Quarto の導入とその活用

西田孝三

地球流体データ解析・数値計算ワークショップ

2023-03-29

## 西田孝三



- kozo2
- バイオインフォマティクスの Research Software Engineer (RSE)
- ?? 年から PyData Osaka オーガナイザー
- 2020 年から Bioconductor の Community Advisory Board メンバー
- 2021 年から The Carpentries のインストラクター
- Ruby アソシエーション開発助成金 2022, 2018, 2016  
(Ruby に詳しいわけではないのですが...)

## 今日の話

- RubyData エコシステムとは何か？
- RubyData エコシステムに  
Quarto という科学技術出版システムを  
導入してみた  
(のでそれについて知ってもらいたい)

# RubyData エコシステム

- Python の SciPy / PyData スタック 等と呼ばれるデータ解析環境の Ruby 版



[http://chris35wills.github.io/courses/pydata\\_stack/](http://chris35wills.github.io/courses/pydata_stack/)  
から引用

## (データサイエンスの) エコシステムを構成するもの

- NumPy 相当 (ベクトル計算、多次元配列、高水準の数学関数...)
- Jupyter Notebook (Mathematica ノートブック のようなもの)
- SciPy (データ解析・数値計算でよく使う関数群)
- matplotlib (汎用可視化ツール)
- ...
- ...

## RubyData で相当する gem

- NumPy -> numo-narray
- Jupyter Notebook -> iruby (Jupyter の「カーネル」)
- SciPy -> ?
- matplotlib -> matplotlib.rb (や charty)
- ...
- ...

## take-home message

Python の(データ解析・数値計算における) 隆盛は、基礎となるパッケージから層積み上げ(スタック)型で依存・拡張することで実現された。

(それをふまえ、私達は Ruby において何ができるか?)

## 自己の活動例

- RubyData エコシステムへの「Quarto」の導入
  - Ruby アソシエーション開発助成金 2022



開発助成 で行ったこと

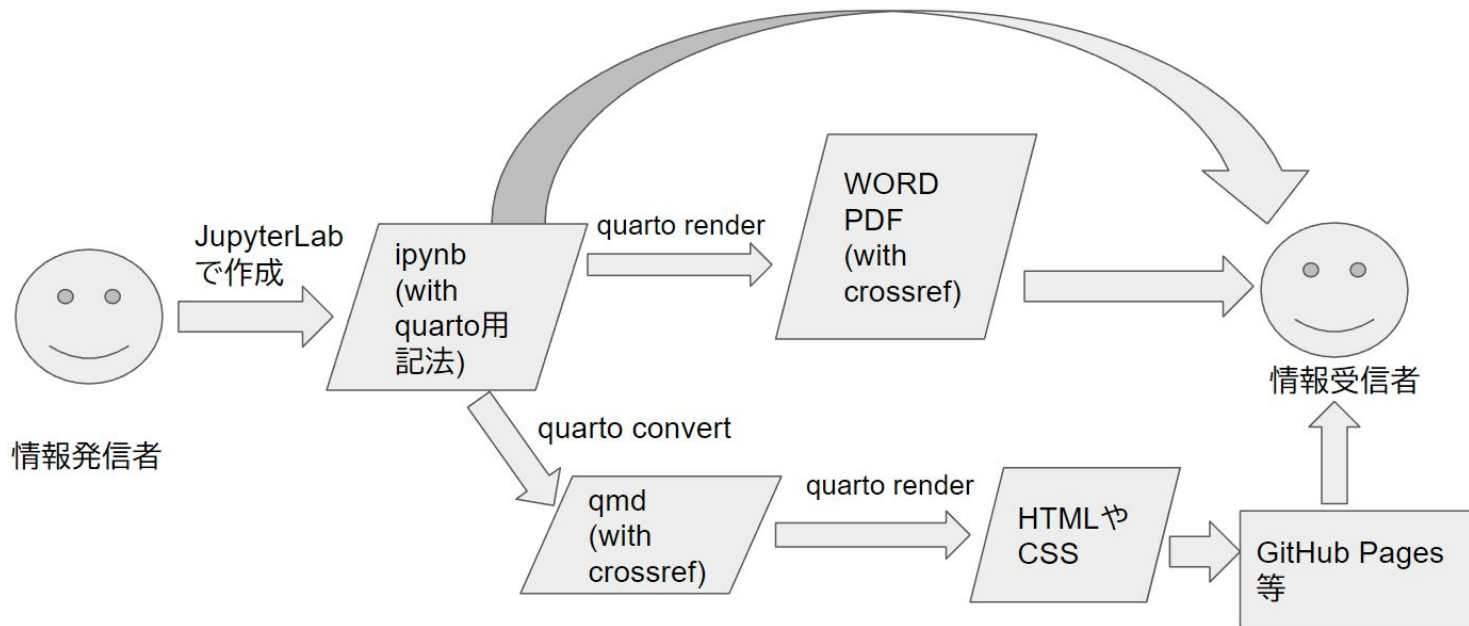
== Quarto が iruby notebook の上にうまく「積み上げる」ことができるか？

の「確認」の「手を動かした」(だけ)



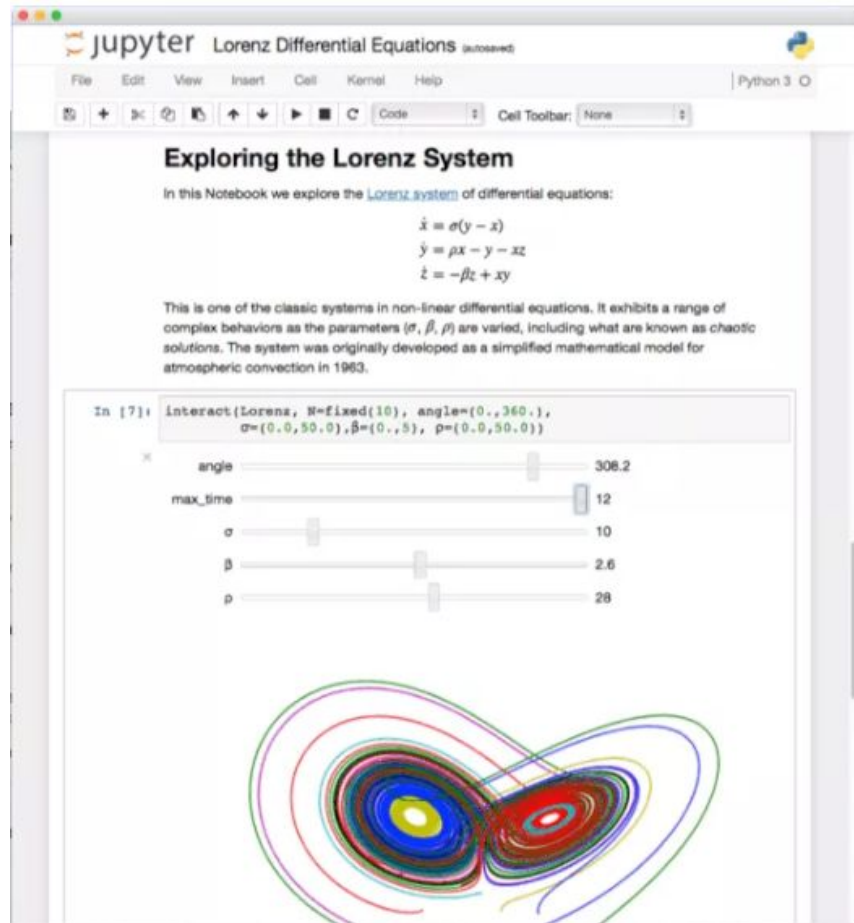
# Quarto とは何か？

- <https://quarto.org> 曰く「An open-source scientific and technical publishing system」
- 私感では「ipynb ファイルから科学技術知識共有のための file convertを行う」もの



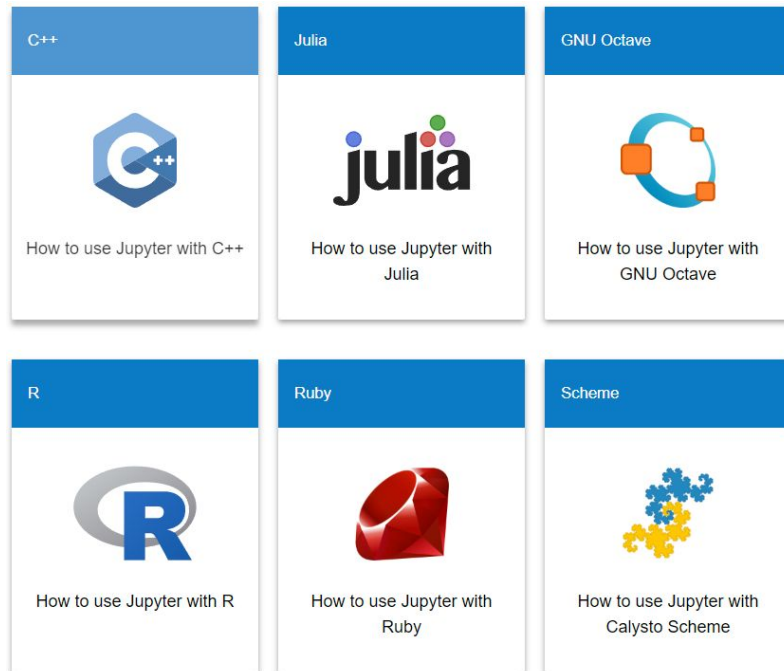
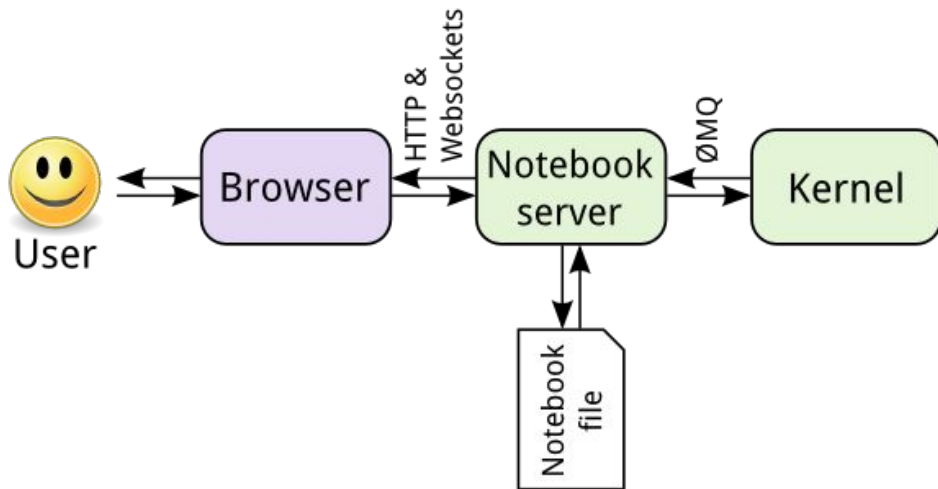
# なぜ Jupyter Notebook を使うのか？

- Webブラウザ上でプログラミングを行う時代だから
  - インタラクティブシェルと画を出す「場所」はターミナル+別窓ではなく、「ブラウザ内」
- 「スタック」の基盤の1つとして盤石になっているから



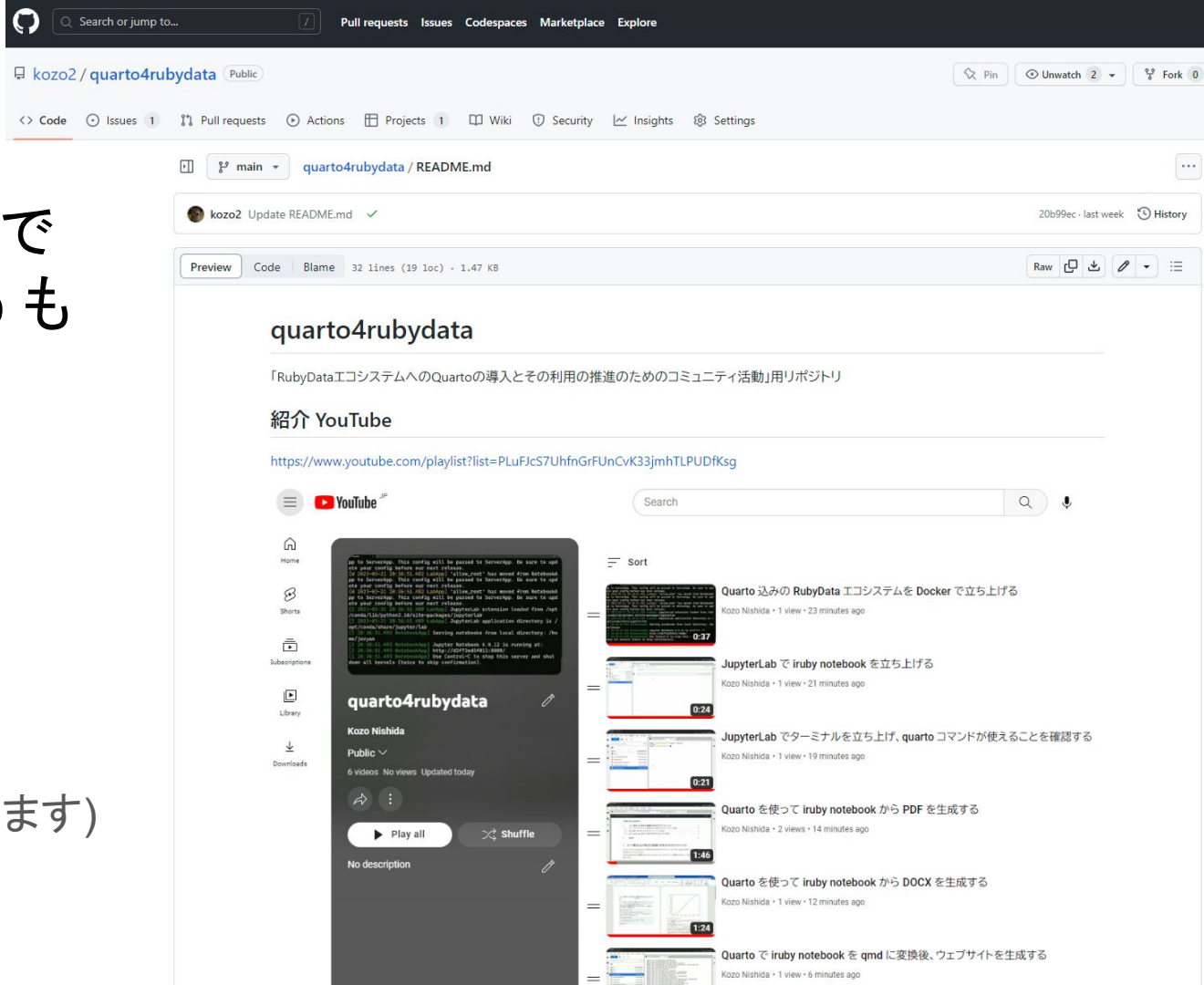
# Ruby も Jupyter Notebook で使える

- Jupyter は Python以外の言語も実行できる設計を持つ



<https://www.pybloggers.com/2016/06/using-remote-kernels-with-jupyter-notebook-server/> から引用

<https://jupyter.org/try>



# Ruby が Jupyter で使える == Quarto も活用できる？

活用できることを「Ruby アソシエーション 開発助成金 2022」で確認した

(のでそれを以降で紹介します)

# 成果例

- だれでもすぐに RubyData+Quarto を試すことができる環境の提供
  - Docker と Binder
- Jupyter (iruby) Notebook [.ipynbファイル]
  - > 科学技術出版用 PDF, DOCX の生成
- Jupyter (iruby) Notebook [.ipynbファイル]
  - > ウェブサイトの生成・デプロイ・公開

なぜ計算環境を提供しなければならないのか？

- ソフトウェアの依存関係が複雑なので、ゼロからの環境の構築は難しい
- データ解析者がデータ解析に集中できるようにしたい

# Docker (コンテナ)とは

```
FROM rubydata/minimal-notebook:latest
ENV TAG=latest

RUN gem install terminal-table

RUN gem update \
  red_arrow\
  red_amber\
  red-datasets\
  red-datasets-arrow\
  red-arrow-numo-narray\
  matplotlib\
  rover-df

Add *.ipynb ./
Add *.png ./
Add *.yml ./

USER root
RUN apt-get update && apt-get install -y --no-install-recommends \
  texlive-lang-japanese \
  texlive-xetex \
  texlive-luatex \
  pandoc \
  pandoc-citeproc \
  lmodern \
  curl \
  zip \
  gdebi-core \
  && rm -rf /var/lib/apt/lists/*

RUN curl -LO https://quarto.org/download/latest/quarto-linux-amd64.deb
RUN gdebi --non-interactive quarto-linux-amd64.deb
RUN rm quarto-linux-amd64.deb
```

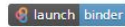
- 仮想マシンと異なり、  
ホストOSの資源を共有利用する仮想化
- Ruby開発助成では先の Dockerfile を利用
- Dockerfile 上部が RubyData 環境用
- Dockerfile 下部が Quarto 環境用
- セットアップの悩みからの開放  
-> データ解析再現性の改善

# Binder とは

- 実演します...



(セットアップ不要で) このプロジェクトを試用するには

 launch binder をクリックしてください。

すると "Binder" という「このリポジトリの内容に応じたJupyterLab環境を構築・提供してくれるオンラインサービス」が立ち上げられます。

それが立ち上がり終わるまで少し時間を要しますが気長にお待ちください。

もし問題無く JupyterLab が立ち上がったなら、左のサイドバーに見える ipynb をダブルクリックし、その内容を実行・確認してみてください。

上記の試用を Docker を用いてご自身のローカル環境で行うには

このリポジトリをクローンし、下記のコマンドを実行してください。

```
docker build -t quarto4rubydata .
docker run -it --rm -p 8888:8888 quarto4rubydata jupyter notebook --NotebookApp.default_url=/lab/ --ip=0.0.0.
```

ウェブブラウザで localhost:8888 を開いてください。

<https://github.com/kozo2/quarto4rubydata> より



# Quarto 込みの RubyData エコシステムを Docker で立ち上げる

- 実演します。(See also <https://www.youtube.com/watch?v=S3dCfEPVpk> )

```
pp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[W 2023-03-21 20:36:51.482 LabApp] 'allow_root' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[W 2023-03-21 20:36:51.482 LabApp] 'allow_root' has moved from NotebookApp to ServerApp. This config will be passed to ServerApp. Be sure to update your config before our next release.
[I 2023-03-21 20:36:51.489 LabApp] JupyterLab extension loaded from /opt/conda/lib/python3.10/site-packages/jupyterlab
[I 2023-03-21 20:36:51.489 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 20:36:51.493 NotebookApp] Serving notebooks from local directory: /home/joyan
[I 20:36:51.493 NotebookApp] Jupyter Notebook 6.4.12 is running at:
[I 20:36:51.493 NotebookApp] http://d2f72ed5f812:8888/
[I 20:36:51.493 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
```

# Quarto を使って iruby notebook から PDF を生成する

- 実演します (See also <https://www.youtube.com/watch?v=wMf02-H9BRk> )



# Quarto を使って iruby notebook から DOCX を生成する

- 実演します。(See also <https://www.youtube.com/watch?v=hDDBjXBCzg0> )



# ジャーナルへの投稿用 Quarto テンプレート



quarto-journals

Overview Repositories 9 Projects Packages People

README.md

## Quarto Journal Templates

The `quarto-journals` organization collects a curated set of journal templates for Quarto.

Use a journal template with the command:

```
quarto use template quarto-journals/<template-name>
```

Journal / Publisher	Name	Install
Association of Computing Machinery	<code>acm</code>	<code>quarto use template quarto-journals/acm</code>
American Chemical Society	<code>acs</code>	<code>quarto use template quarto-journals/acs</code>
Biophysical journal	<code>biophysical-journal</code>	<code>quarto use template quarto-journals/biophysical-journal</code>
Elsevier Journals	<code>elsevier</code>	<code>quarto use template quarto-journals/elsevier</code>
American Statistical Association Journals	<code>jasa</code>	<code>quarto use template quarto-journals/jasa</code>
Journal of Statistical Software	<code>jss</code>	<code>quarto use template quarto-journals/jss</code>
Public Library of Science	<code>plos</code>	<code>quarto use template quarto-journals/plos</code>

### Using a Template

These templates require features in Quarto 1.2, so you will currently need [pre-release versions of Quarto](#).

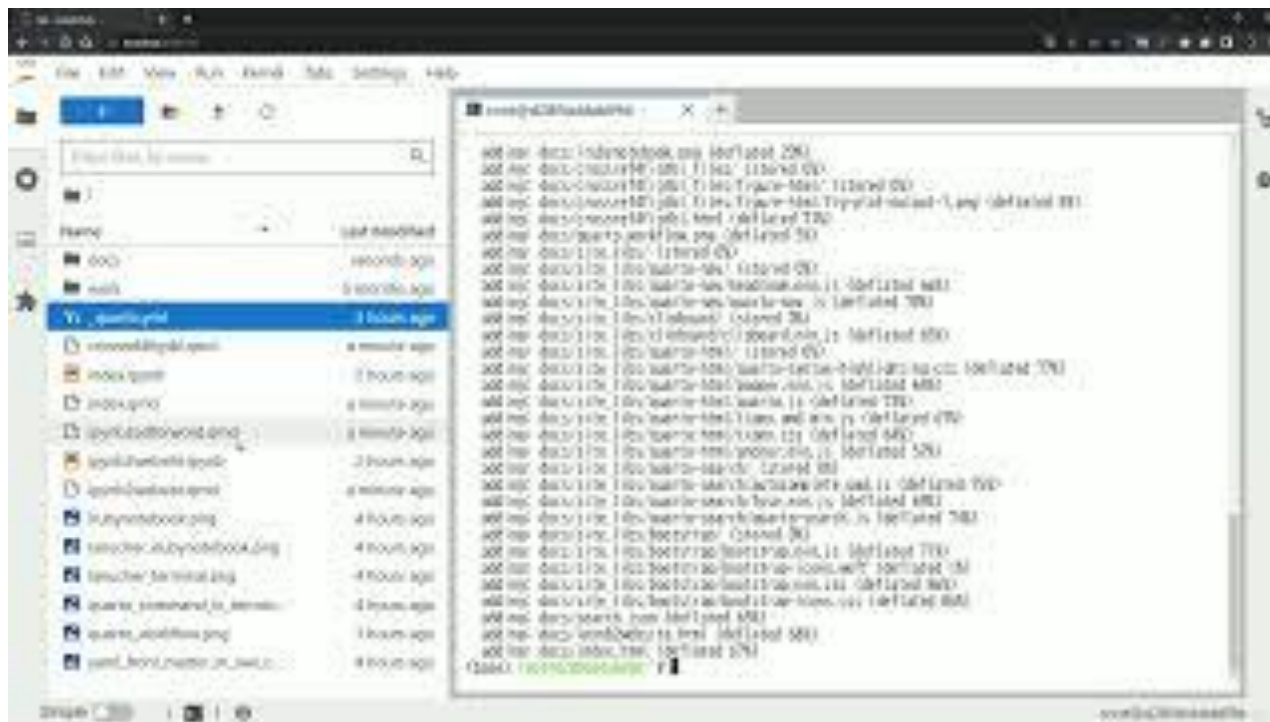
- (ipynb ではなく) qmd の  
ジャーナル投稿用テンプレートがある

- <https://github.com/quarto-journals>

- <https://quarto.org/docs/journals/>

# Quarto で iruby notebook を qmd に変換後、ウェブサイトを生成する

- 実演します。(See also <https://www.youtube.com/watch?v=Sw26N5UggRQ> )



# Quarto 利用によって得られるもののまとめ

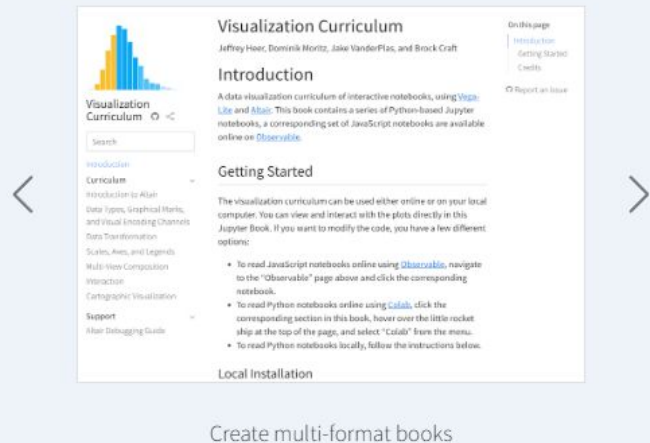


Overview Get Started Guide Extensions Reference Gallery Blog Help ▾

## Gallery

Quarto can produce a wide variety of output formats. Here are some examples:

- [Articles & Reports](#)
- [Presentations](#)
- [Interactive Docs](#)
- [Websites](#)
- [Books](#)



<https://quarto.org/docs/gallery/>

# 「地球流体データ解析・数値計算」 でRubyData/Quarto を役立てるには？

- 何はともあれ  
Jupyter (iruby) Notebook で  
記事を書く！
  - matplotlib.rb 等を用いて  
「地図投影」「流れ場の可視化」を行う？
- その後の知識共有をQuartoが  
役立つかもしれません！

Pythonによる気象データ  
サイエンス

この本を検索...

Pythonによる気象データサイエンス

はじめの一步

様々なデータの出力と可視化

高層気象観測データ

アメダスと天気予報

海面水温

気象衛星ひまわり

再解析データ

数値予報データ

力学解析

統計解析

参考文献

謝辞

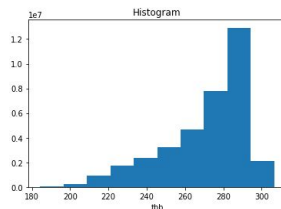
Python環境の構築

Powered by Jupyter Book

```
import matplotlib.pyplot as plt
import xarray as xr

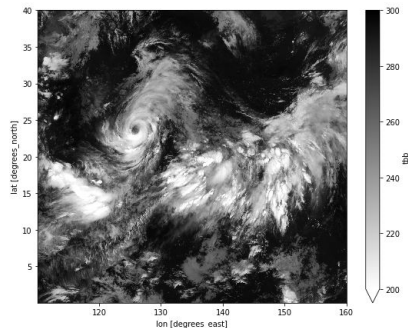
xr_tbb = xr.DataArray(np.float32(dataTBB), name="tbb",
                      coords = {
                        'lat': ('lat', lat, {'units': 'degrees_north'}),
                        'lon': ('lon', lon, {'units': 'degrees_east'}),
                      },
                      dims = ['lat', 'lon'])

xr_tbb.plot.hist()
plt.show()
```



$T_{BB}$ の範囲は200から300で良さそうです。

```
xr_tbb.loc[48:0, 118:168].plot.imshow(vmin=288, vmax=308,
                                       interpolation="none", cmap="gray_r", figsize=[9,7])
plt.show()
```



Previous  
海面水温

Next  
再解析データ

著者 榎本剛  
© Copyright 2021.

Pythonによる気象データサイエンス

<https://www.dpac.dpri.kyoto-u.ac.jp/enomoto/pymetds/index.html>

のスクリーンショット