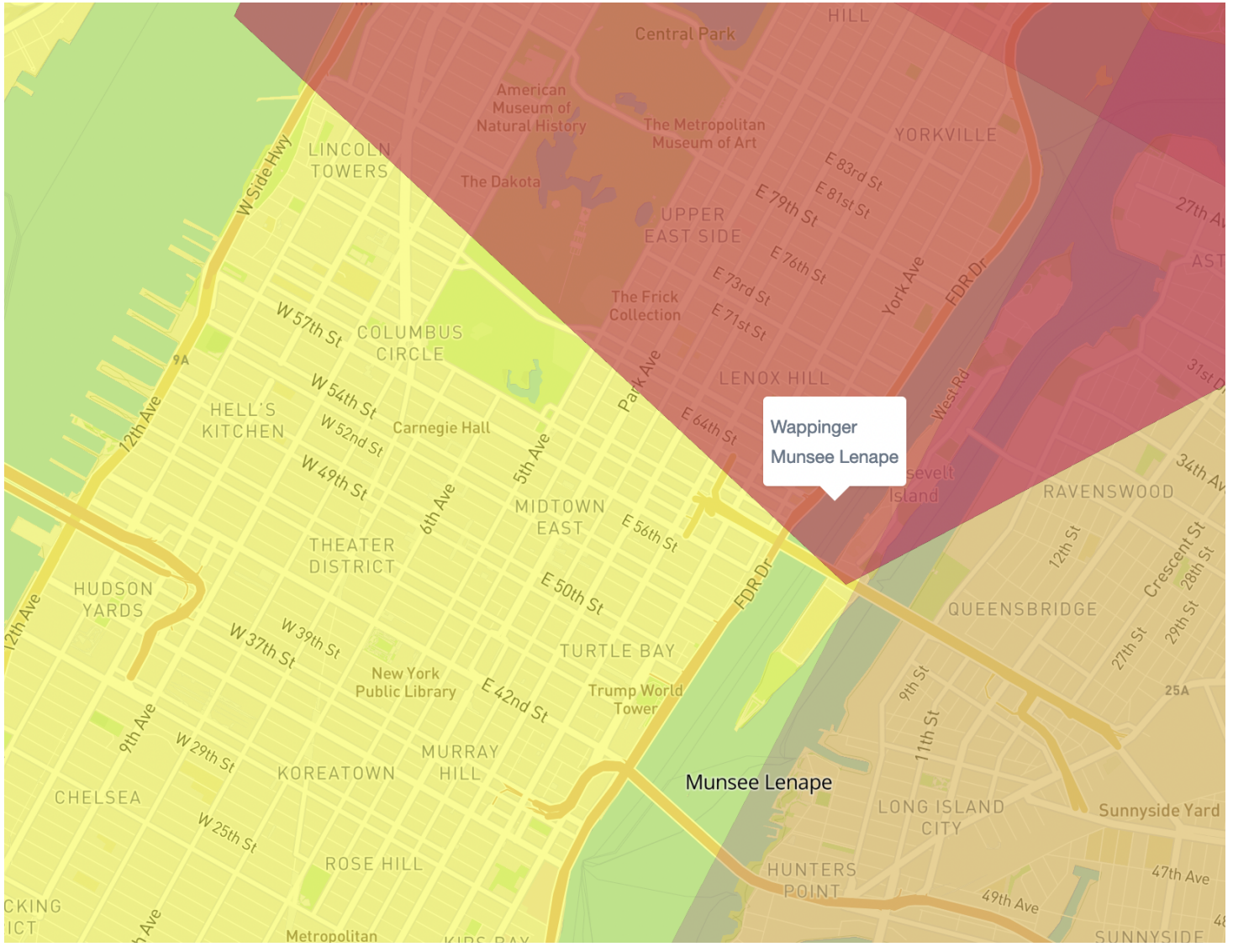
Moving to Quarto from RMarkdown and Python Jupyter Notebooks

NYR Conference 2023

Daniel Chen

Hello

Munsee Lenape



Daniel Chen

@chendaniely

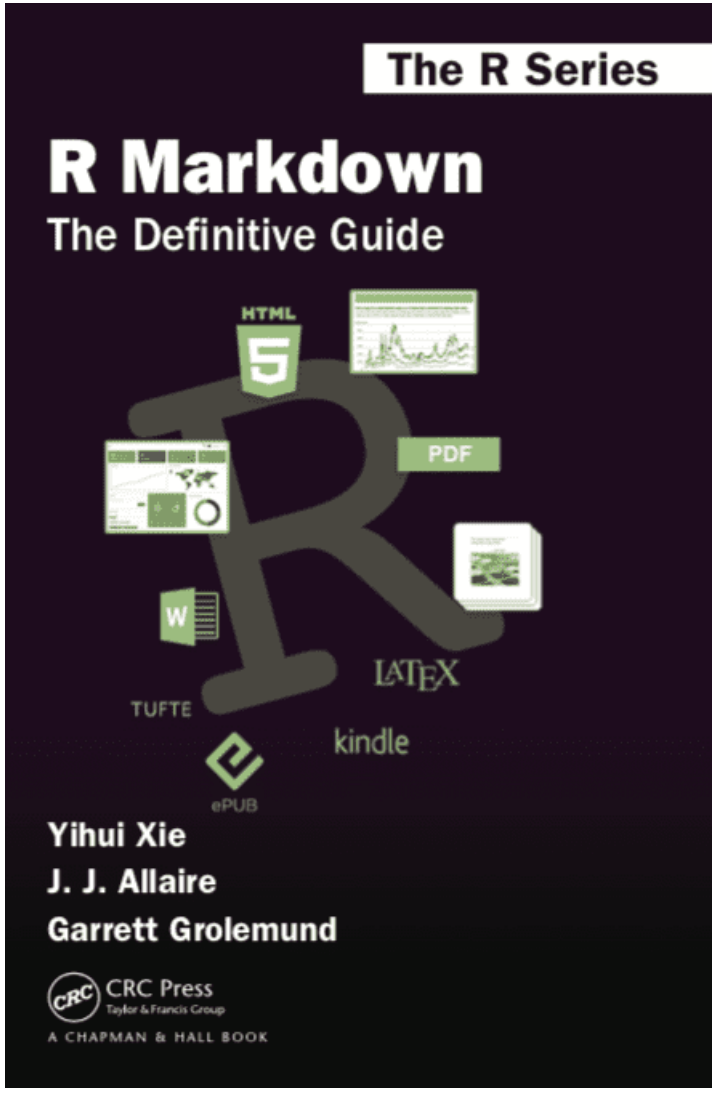
- Postdoctoral Research and Teaching Fellow
  UBC, MDS-Vancouver
- Data Science Educator,
  Posit, PBC
- The Carpentries
- Author, Pandas for Everyone

Literate Programming

Why Literate Programming?

- Data Scientist
  - RMarkdown + Jupyter Notebooks
    - Analysis
    - Reports + Documentation
- Academic
  - Papers
- Technical Writer
  - Blog
  - Website
  - Presentation
  - Book

RMarkdown


The R Series
R Markdown
The Definitive Guide
HTML
PDF
W
LATEX
TUFTE
kindle
ePUB
Yihui Xie
J. J. Allaire
Garrett Grolemund
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK


rmarkdown

Code Chunks

````
```{r}
cmv <- read_excel("data/cmv.xlsx")
head(cmv)
```
````

RMarkdown Document

````
---
title: "example-analysis"
author: "Daniel Chen"
output: html_document
---
```{r setup, include=FALSE}
library(tidyverse)
library(readxl)
library(writexl)
```
## Load Data
Retrospective Cohort Study of the Effects of
Donor KIR genotype on the reactivation of cytomegalovirus (CMV)
after myeloablative allogeneic hematopoietic stem cell transplant.
```{r}
cmv <- read_excel("data/cmv.xlsx")
head(cmv)
```
````

Render `.Rmd` with `{rmarkdown}`

Demo file: `example-analysis.Rmd`

Render Command:

```
Rscript -e "rmarkdown::render('example-analysis.Rmd')"
```

Specify output file (and location):

```
Rscript -e "rmarkdown::render(
    input = 'example-analysis.Rmd',
    output_file = 'output/010-example-analysis-rmd.html')"
```

Render `.Rmd` with `quarto`

Demo file: `example-analysis.Rmd`

Render Command:

```
quarto render example-analysis.Rmd
```

Specify output file:

```
# output folders only work with quarto projects
touch _quarto.yml

quarto render example-analysis.Rmd \
    --toc \
    --output output/020-example-analysis-rmd-qmd.html
```

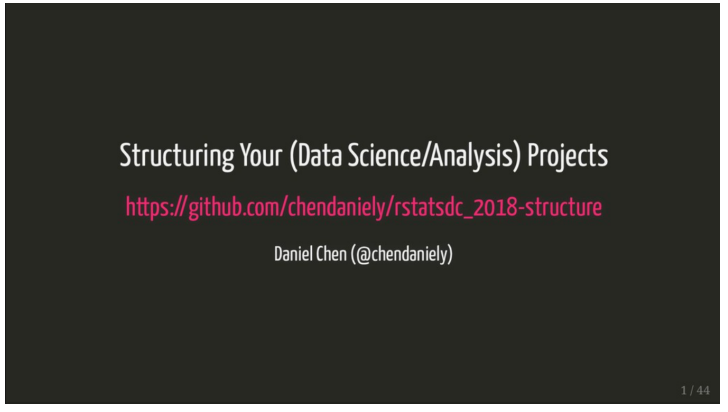
- `quarto` is command line tool!

Caveat: Single Quarto Document

- Output directory Github Discussion
  - https://github.com/quarto-dev/quarto-cli/discussions/2171
- Pre and Post Render
  - https://quarto.org/docs/projects/scripts.html#pre-and-post-render

Project templates

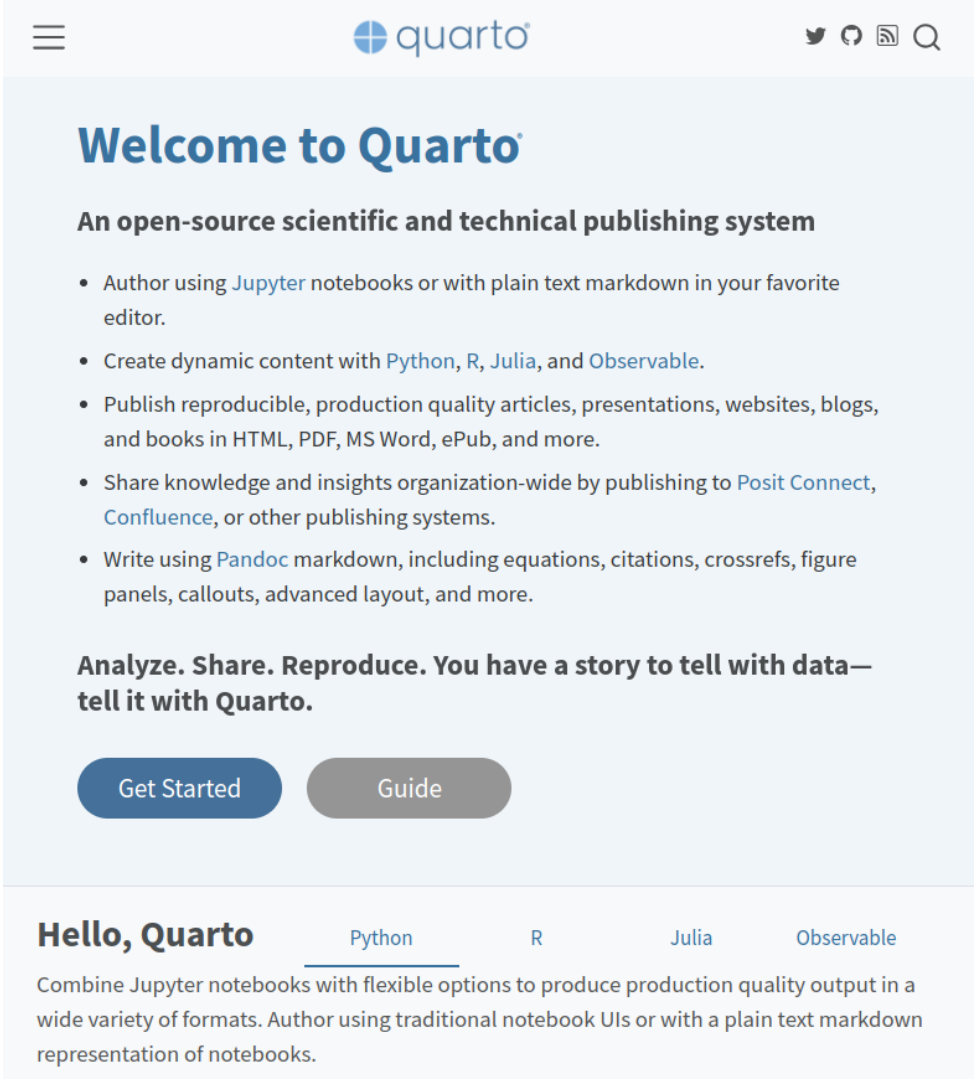
DCR 2018: Structuring Your (Data Science/Analysis) Projects



NYR 2019: Building Reproducible and Replicable Projects

Tiffany Timbers DSCI 310: Reproducible and trustworthy workflows for data science: https://ubc-dsci.github.io/dsci-310-student/

Quarto



https://quarto.org/

- Plain text source document
- Literate programming
- Multiple language support
  - Even in the same document!
- Multiple output formats
  - Pandoc + Markdown
- Familiar

- Quarto Gallery: https://quarto.org/docs/gallery/
- Quarto Guide: https://quarto.org/docs/guide/
- Quarto Reference: https://quarto.org/docs/reference/

Quarto Documents

RMarkdown YAML

```yaml
---
title: "Example Analysis"
subtitle: "RMarkdown"
author: "Daniel Chen"
output: html_document
---
```

Quarto YAML

```yaml
---
title: "Example Analysis"
subtitle: "Quarto"
author: "Daniel Chen"
format: html
---
```

RMarkdown and Quarto chunk options:

````
```{r setup}
#| include: false
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(readxl)
library(writexl)
```
````

Render a Quarto Document

Demo file: `example-analysis.qmd`

Render Command:

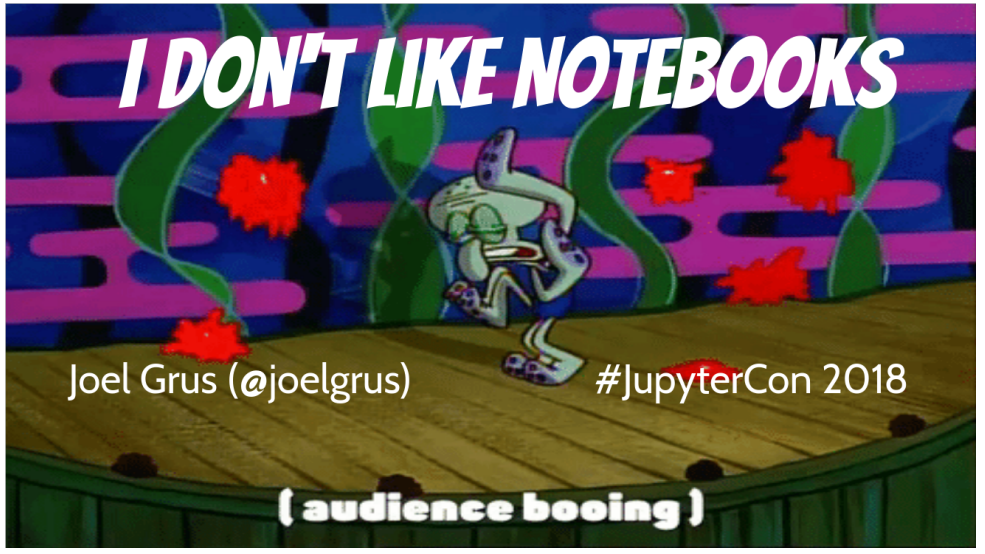
```
quarto render example-analysis.qmd
```

Specify output file:

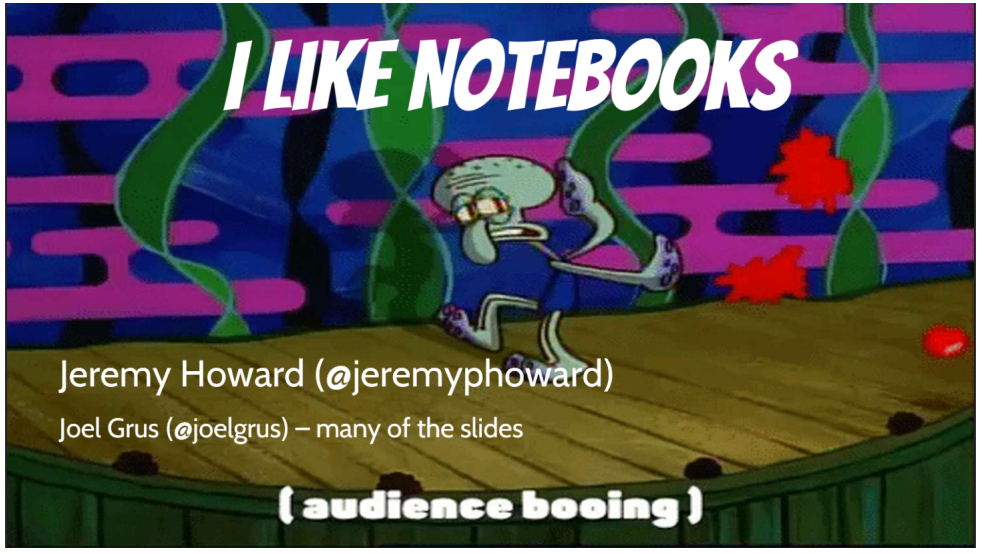
```
quarto render example-analysis.Rmd \
    --toc \
    --output-dir output \
    --output 030-example-analysis-rmd-qmd.html
```

Jupyter

Notebooks



Youtube



Youtube

Daniel's List

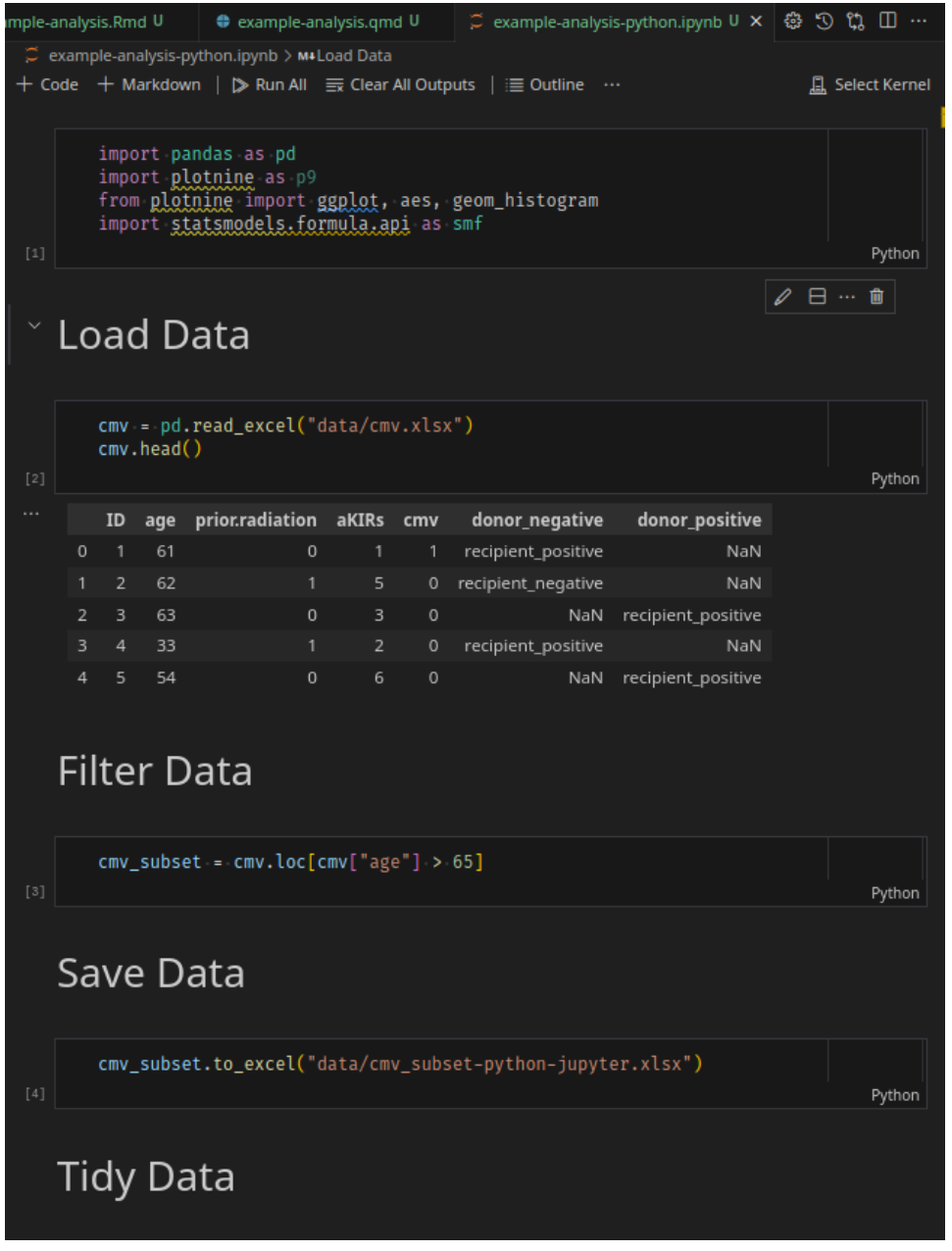
- Technical Writing
  - ✅ Literate programming
  - ❌ Editing JSON
- Data Science
  - More an output format than a source document
  - ✅ Great for posting code+output (e.g. a workshop)
  - ❌ Not great for source control collaborative document
- Teaching
  - ✅ `nbgrader` for course assignment creation + grading
  - ✅ Restart Kernel > Run All

Jupyter Notebooks are JSON

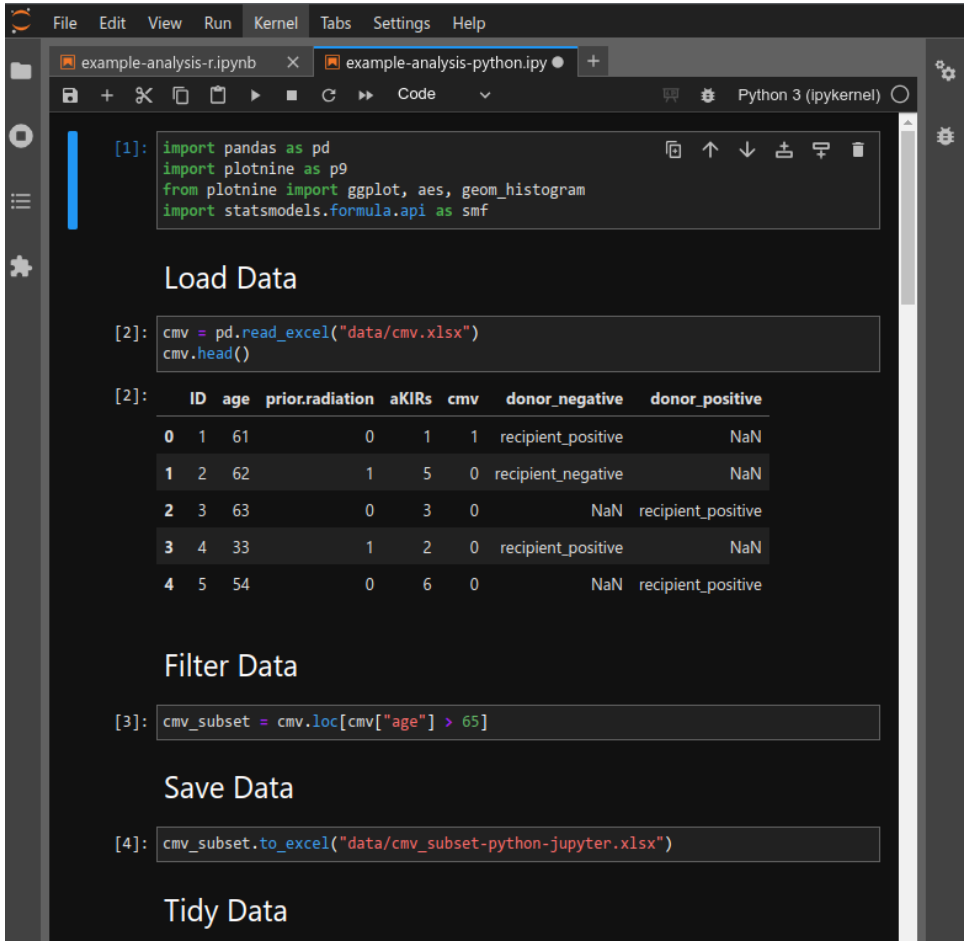
```
{
 "cells": [
  {
   "cell_type": "code",
   "execution_count": 1,
   "id": "4a9a7246-de20-4aac-945a-b8f0e7db0ac6",
   "metadata": {},
   "outputs": [],
   "source": [
    "import pandas as pd\n",
    "import plotnine as p9\n",
    "from plotnine import ggplot, aes, geom_histogram\n",
    "import statsmodels.formula.api as smf"
   ]
  },
  {
   "cell_type": "markdown",
   "id": "8f8205a7-a172-492a-bb22-e24bc1fc7ce2",
```

Need Something to View + Render

VSCode



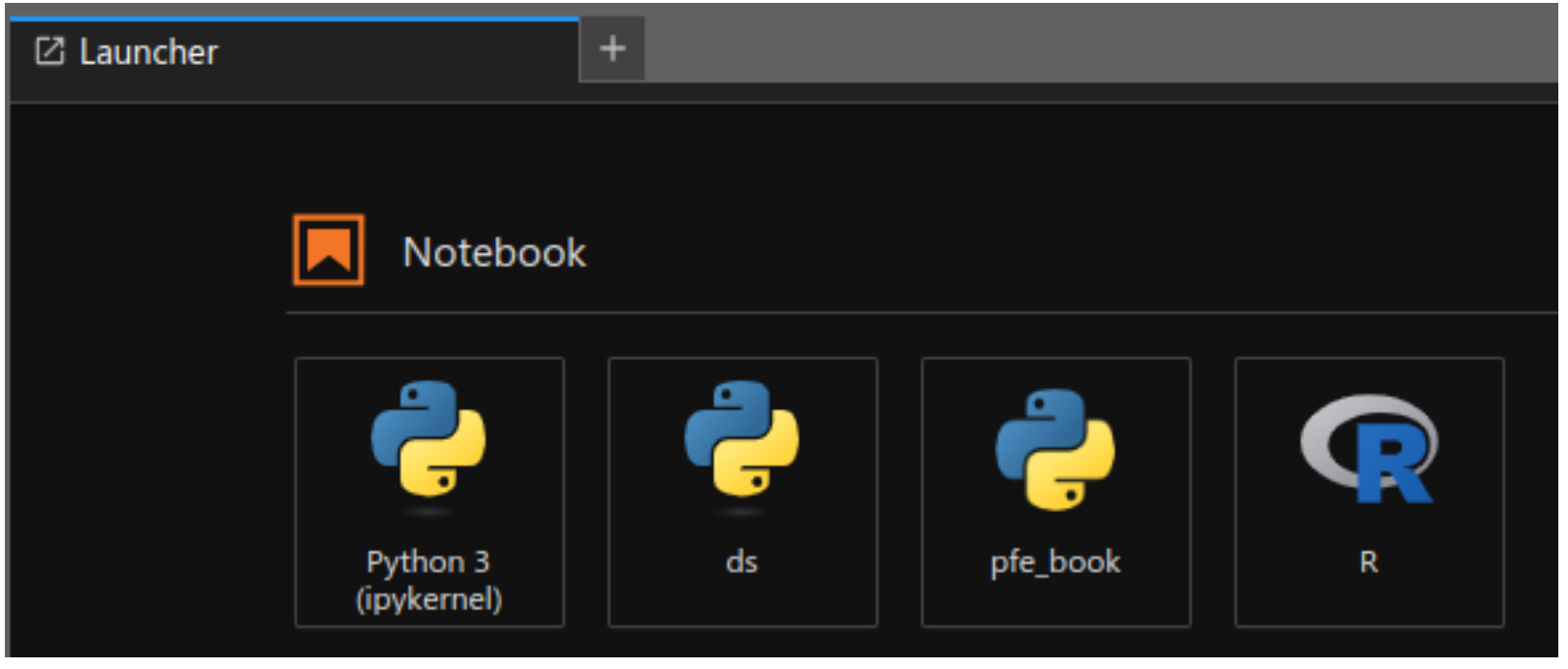
Jupyter Lab



Jupyter does R!

- You need the `IRKernel` installed: https://github.com/IRkernel/IRkernel

```
1 install.packages('IRkernel')
2 IRkernel::installspec()
```



Render `.ipynb` with `nbconvert`

- Demo files:
  - `example-analysis-python.ipynb`
  - `example-analysis-r.ipynb`

Python Kernel:

```
jupyter nbconvert \
    --to html \
    --output output/040-example-analysis-python-jupyter.html \
    --execute example-analysis-python.ipynb
```

R Kernel:

```
jupyter nbconvert \
    --to html \
    --output output/050-example-analysis-r-jupyter.html \
    --execute example-analysis-r.ipynb
```

(Hint: they're the same command)

Jupyter Notebook as a Source Document

To make your version control diffing easier, you may want to clear the output from the notebook JSON file.

In `nbconvert 6.0+`, you can use `--clear-output --inplace`:

```
jupyter nbconvert --clear-output --inplace example-analysis-python.ipynb
jupyter nbconvert --clear-output --inplace example-analysis-r.ipynb
```

Or use the `--to notebook` argument if you want to preserve a rendered notebook

Render `.ipynb` with `quarto`

Takes whatever is in the notebook (no additional execution) and rendered (to html by default)

```
quarto render example-analysis-python.ipynb
quarto render example-analysis-r.ipynb
```

Use `--execute` to execute the cells and render

```
quarto render example-analysis-python.ipynb --execute
quarto render example-analysis-r.ipynb --execute
```

Render `.ipynb` with `quarto`

Python Kernel:

```
quarto render example-analysis-python.ipynb \
    --to html \
    --execute \
    --toc \
    --output-dir output \
    --output 060-example-analysis-python-ipynb.html
```

R Kernel:

```
quarto render example-analysis-r.ipynb \
    --to html \
    --execute \
    --toc \
    --output-dir output \
    --output 060-example-analysis-r-ipynb.html
```

Embed Jupyter output in Quarto

From a Jupyter notebook with code output:

- Demo files:
  - `example-analysis-python-qmd_meta.ipynb`
  - `example-analysis-python-qmd_meta.qmd`

Using a notebook with existing output:

```
jupyter nbconvert \
    --to notebook \
    --execute \
    --inplace \
    example-analysis-python-qmd_meta.ipynb
```

You can add quarto `#|` metadata comments to a cell, and use jupyter output directly in a quarto document

Embed Jupyter output in Quarto

```
#| label: fig-age_hist
#| fig-cap: >
#|     A histogram of the ages in our Cytomegalovirus dataset
ggplot(cmv_tidy, aes(x="age")) + geom_histogram()
```

Use a quarto shortcode:

```
{{< embed example-analysis-python-qmd_meta.ipynb#fig-age_hist >}}
```

Render the example:

```
quarto render example-analysis-python-qmd_meta.qmd \
    --to html \
    --output-dir output \
    --output 080-example-analysis-python-qmd_meta.html
```

https://quarto.org/docs/authoring/notebook-embed.html

Converting

jupytext

https://jupytext.readthedocs.io/

Rmd -> qmd

```
jupytext \
    --to qmd \
    --output output/090-convert-rmd_qmd.qmd \
    example-analysis.Rmd
```

ipynb -> qmd

```
jupytext \
    --to qmd \
    --output output/100-convert-ipynb_qmd.qmd \
    example-analysis-python.ipynb
```

quarto convert

```
quarto convert example-analysis-python.ipynb \
    --output output/120-convert-ipynb_qmd.qmd
```

Publication

Publish your files

```
quarto publish              # Publish Project (ask provider)
quarto pubish talk.qmd      # Publish document (ask provider)

quarto publish quarto-pub   # Quarto.pub

quarto publish gh-pages     # GitHub Pages
quarto publish netlify      # Netlify

quarto publish connect      # RStudio Connect
quarto publish confluence   # Confluence
```

https://quartopub.com/

Thanks!

Thanks

@chendaniely

- github.com/chendaniely/rstatsnyc-2023-quarto
- chendaniely.quarto.pub/rstatsnyc-rmd-jupyter-quarto/