

The right tool for the job

SciPy 2024 | Julia Silge



Hello!



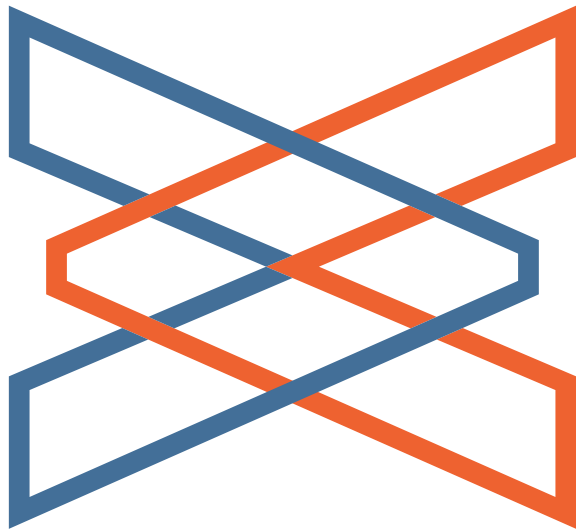
 @juliasilge

 @juliasilge@fosstodon.org

 youtube.com/juliasilge

 juliasilge.com

<https://juliasilge.github.io/scipy2024>



positTM

Tools for data science

- Quarto
- Shiny
- Great Tables
- Vetiver
- Pins
- Exciting new work! 🎉



Using multiple programming languages

- What does it ***cost***?
- What do you ***gain***?
- What can we ***give***?





For the individual

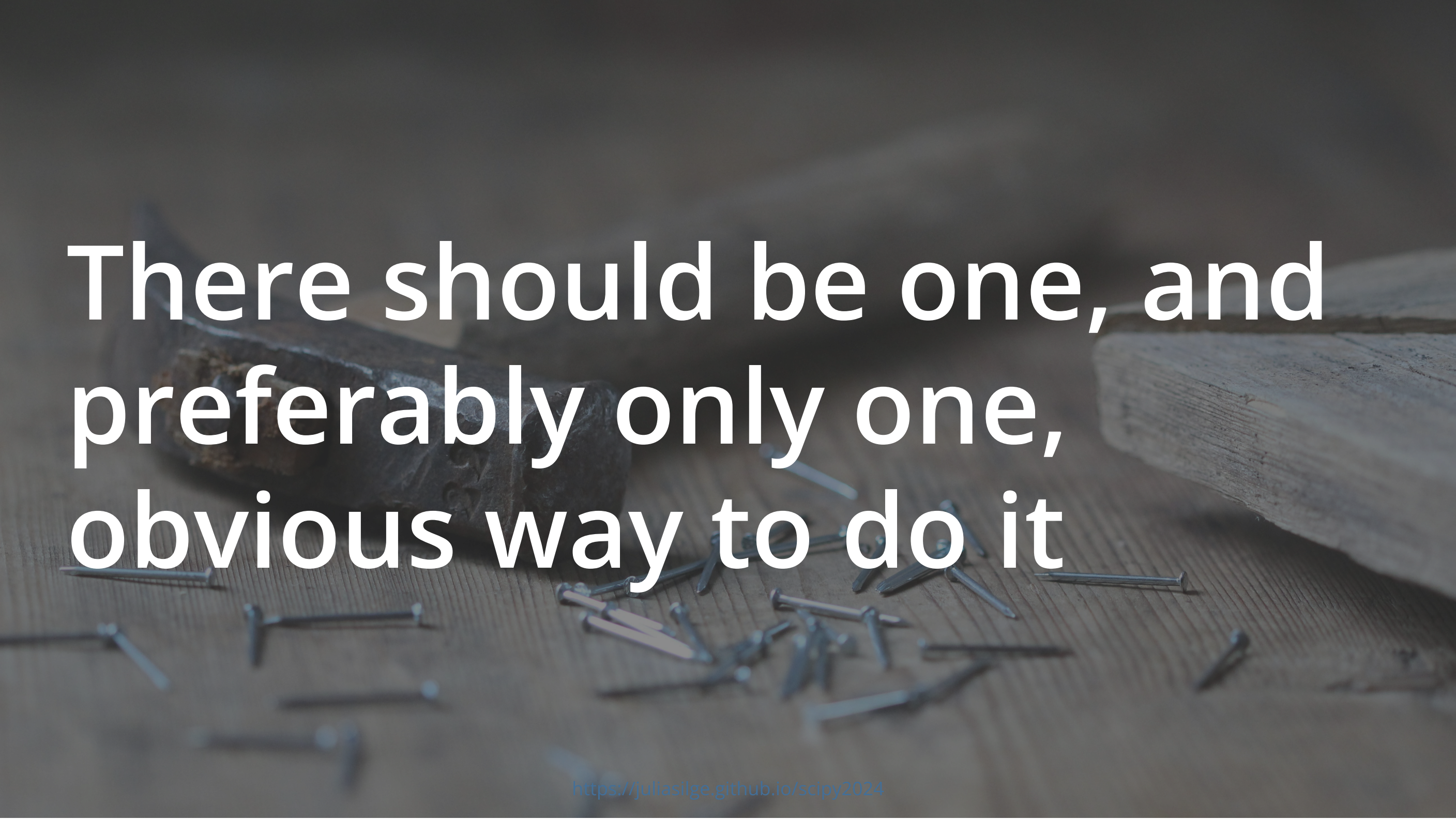
- It is expensive to learn new things
- There are benefits to specialization



In an organization

- Consistency
- Complexity



The background is a dark, moody photograph of a wooden workbench. Numerous metal nails are scattered across the surface, some lying flat and others standing upright. A piece of light-colored wood is visible on the right side, and a dark, possibly charred or stained piece of wood is on the left. The overall lighting is low, creating a sense of depth and texture.

There should be one, and
preferably only one,
obvious way to do it

Pins

Python

```
import pins

board = pins.board_temp()
board.pin_write(
    very_nice_data,
    "important-stuff",
    type = "parquet")
```

R

```
library(pins)

board <- board_temp()
board |> pin_write(
    very_nice_data,
    "important-stuff",
    type = "parquet")
```

Pins

Python

```
import pins

board = pins.board_temp()
board.pin_read("important-stuff")
```

R

```
library(pins)

board <- board_temp()
board |> pin_read("important-stuff")
```

- Cost for individuals
- Cost for our organization



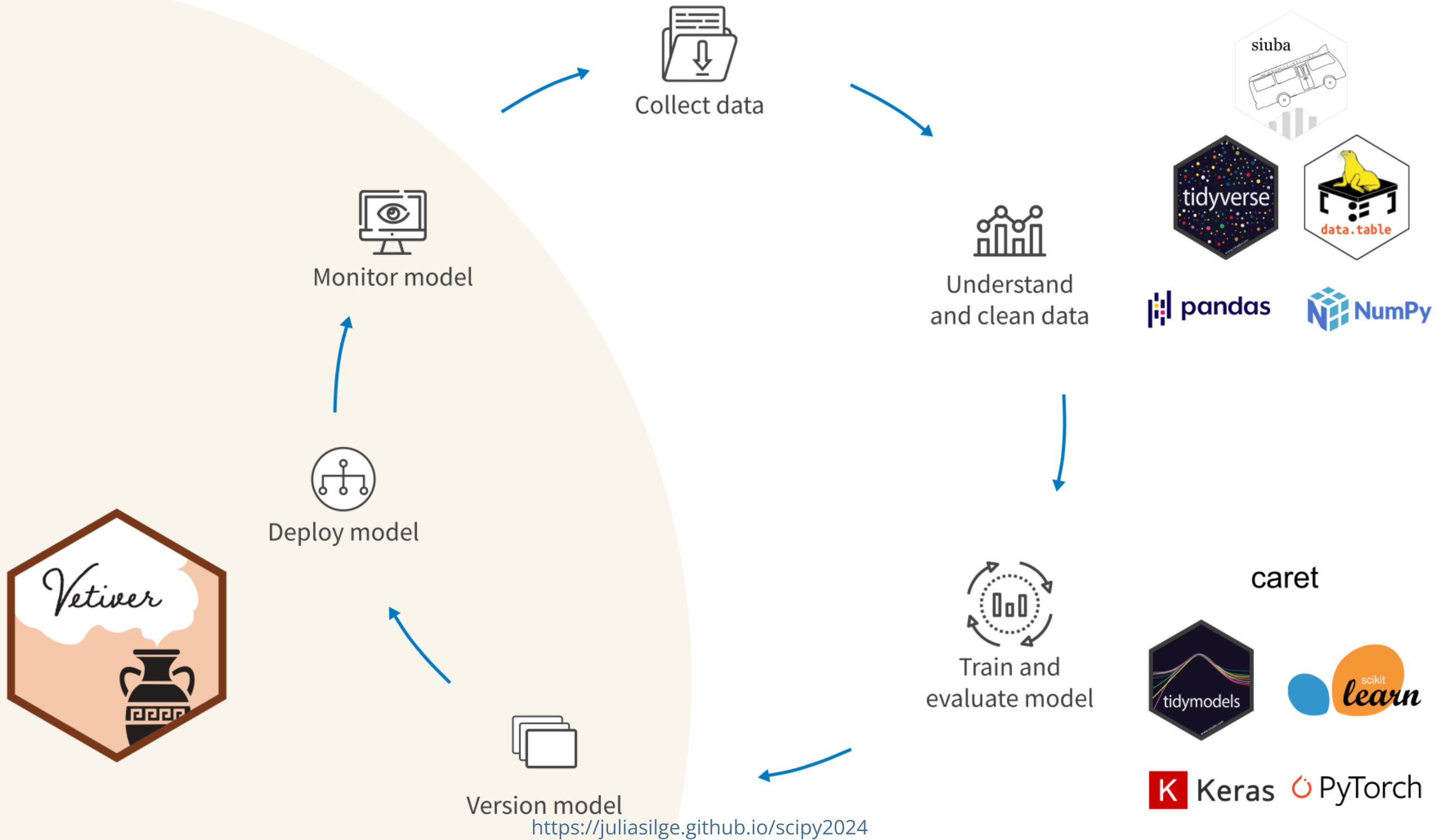
In an organization

- Everyone can be more productive



A dark, curved metal object, possibly a piece of a sword or a tool, lies on a wooden surface. Scattered around it are numerous small, silver-colored nails. The background is a blurred wooden surface, suggesting a workshop or a construction site. The overall tone is dark and moody.

Practicality beats purity



Vetiver

Python

```
from vetiver import VetiverModel, VetiverAPI

v = VetiverModel(
    model_fit,
    "my-important-model",
    prototype_data = X_train)

api = VetiverAPI(v)
api.run()
```




R

```
library(vetiver)
library(plumber)

v <- vetiver_model(
  model_fit,
  "my-important-model")

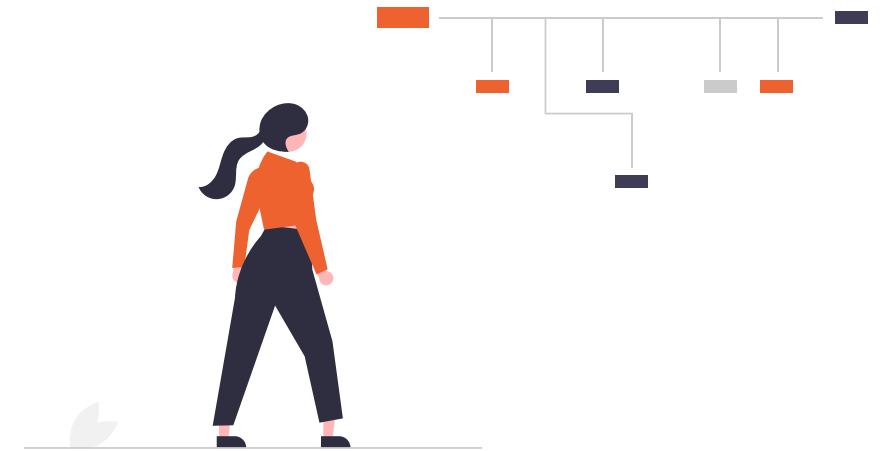
pr() |>
  vetiver_api(v) |>
  pr_run()
```

MLOps is...

- *Versioning*
 - Managing change in models 
- *Deploying*
 - Putting models in REST APIs 
- *Monitoring*
 - Tracking model performance 

For the individual

- You can scale your impact
- Consider the long term
- Increase your vocabulary





Building tools

- Learn from one community
- Bring to a different one





Brooke Rabe

@brookerabe

Question: is there an ide for python as effective as Rstudio for R? For data wrangling, data visualization, and reproducibility via rmarkdown, can python beat R?

11:44 AM · Feb 27, 2024 · 324 Views



Sarah Harris

@implausible17



Python is growing on me as a language - but I still prefer RStudio as IDE - can anyone who uses both recommend a nice Python IDE? In particular one that autosuggests variable names etc as well as RStudio does? & search/filter dataframes? #Rstats



Mike Konczal

@mtkonczal

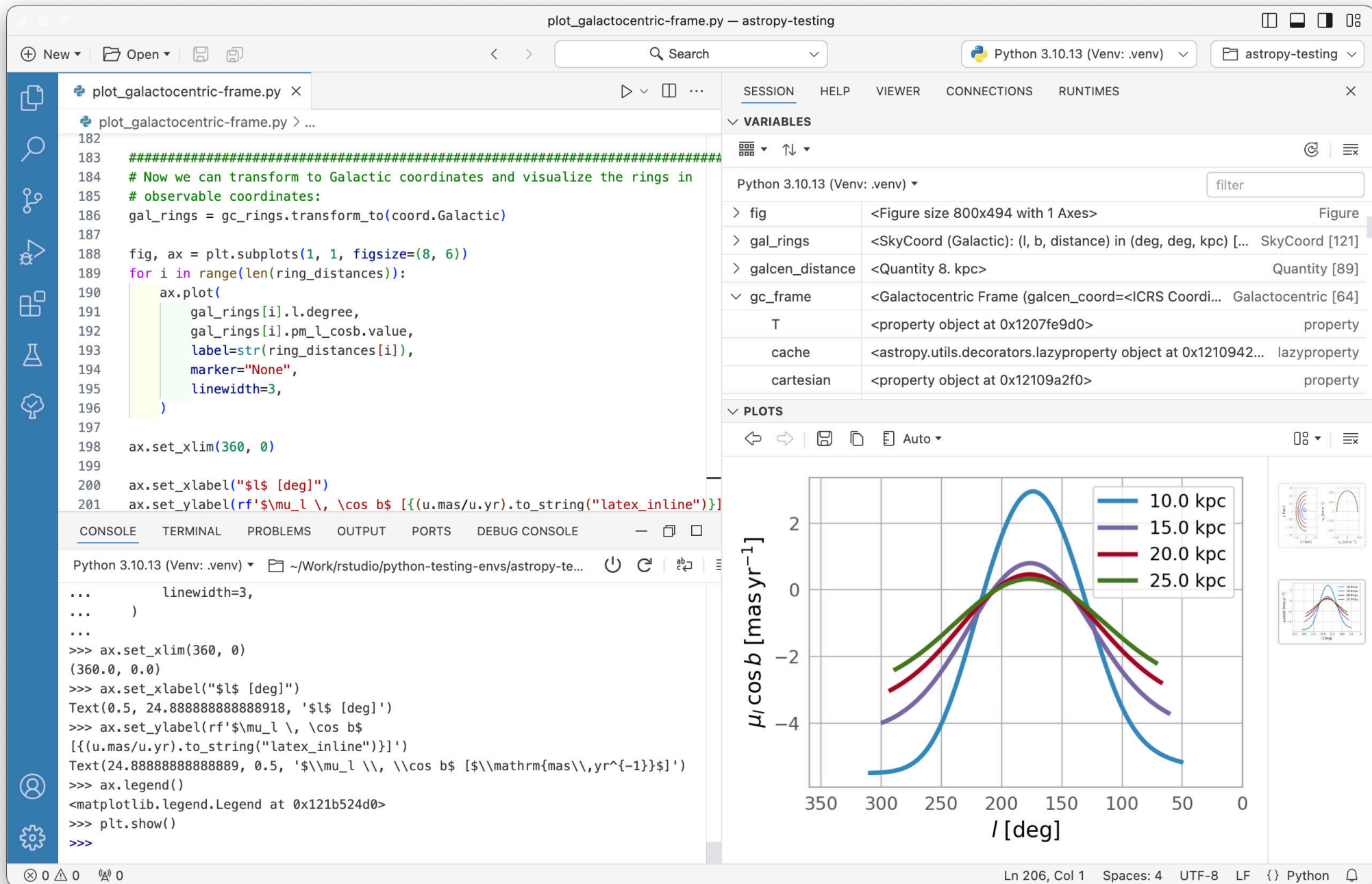
For the Python and R data people out there, I'm looking for a Python IDE environment to RStudio's level (not just data frame/

Trying to do some new data work in Python after a lifetime in R and....how do you look at things? (Realizing how spoiled I am by RStudio.)

10:47 PM · Oct 3, 2023 · 30.6K Views



<https://juliasilge.github.io/scipy2024>



quarto

penguins.qmd — seaborn-testing

New Open Search Python 3.10.13 (Venv: .venv) seaborn-testing

penguins.qmd

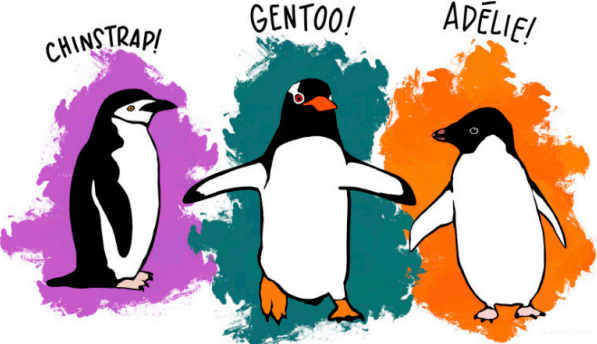
```
1 ---
2 title: "Meet the Palmer penguins!"
3 format: html
4 ---
5
6 The Palmer penguins data contains size measurements for three penguin species
7 observed on three islands in the Palmer Archipelago, Antarctica.
8
9 ![[The Palmer Archipelago penguins. Artwork by @allison_horst.](https://raw.
10 githubusercontent.com/allisonhorst/palmerpenguins/main/man/figures/lter_penguins.
11 png)]
12
13 ::: {.callout-note}
14 That's right -- [Gentoo Linux is named after penguins](https://wiki.gentoo.org/
15 wiki/FAQ)!
16 :::
17
18 Run Cell | Run Next Cell
19 ```{python}
20 import seaborn as sns
21 df = sns.load_dataset("penguins")
22 ```
23
24 These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the [Palmer
25 Station Long Term Ecological Research Program](https://pal.lternet.edu/), part of
26 the [US Long Term Ecological Research Network](https://lternet.edu/). The data
27 were imported directly from the [Environmental Data Initiative](https://
28 environmentaldatainitiative.org/) (EDI) Data Portal, and are available for use by
29 CC0 license ("No Rights Reserved") in accordance with the [Palmer Station Data
30 Policy](https://pal.lternet.edu/data/policies).
31
32 Run Cell | Run Above
33 ```{python}
34 sns.lmplot(
35     data=df,
36     x="bill_length_mm", y="bill_depth_mm", hue="species",
37     height=5
38 )
39 ```
```

SESSION TERMINAL HELP VIEWER CONNECTIONS RUNTIMES

← → ↺ http://localhost:3547/ Zoom: (Auto) ↗

Meet the Palmer penguins!

The Palmer penguins data contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.



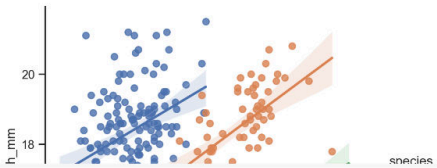
CHINSTRAP! GENTOO! ADÉLIE!

The Palmer Archipelago penguins. Artwork by @allison_horst.

Note
That's right -- [Gentoo Linux is named after penguins](https://wiki.gentoo.org/wiki/FAQ)!

```
import seaborn as sns
df = sns.load_dataset("penguins")
```

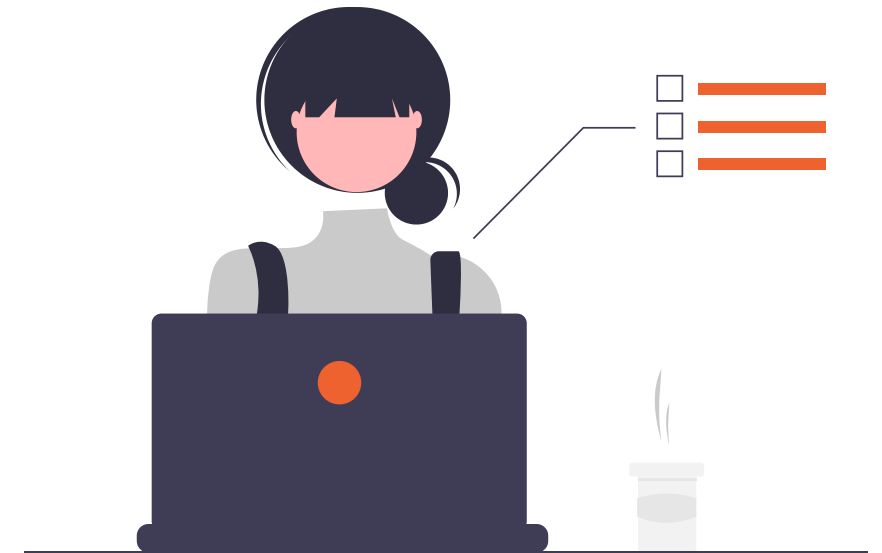
These data were collected from 2007 - 2009 by Dr. Kristen Gorman with the [Palmer Station Long Term Ecological Research Program](https://pal.lternet.edu/), part of the [US Long Term Ecological Research Network](https://lternet.edu/). The data were imported directly from the [Environmental Data Initiative](https://environmentaldatainitiative.org/) (EDI) Data Portal, and are available for use by CC0 license ("No Rights Reserved") in accordance with the [Palmer Station Data Policy](https://pal.lternet.edu/data/policies).



Ln 11, Col 90 Spaces: 4 UTF-8 LF Quarto

Positron

- Positron is a next-generation data science IDE
- Positron is a *very* early stage project
- <https://github.com/posit-dev/positron/>



Thank you!



 @juliasilge

 @juliasilge@fosstodon.org

 youtube.com/juliasilge

 juliasilge.com

<https://juliasilge.github.io/scipy2024>