

Polars: laying the foundation
of the pharamaverse-py

- PhD in Cognitive Psychology
- Open source data tools at Posit, PBC
(e.g. Great Tables)



Bandit

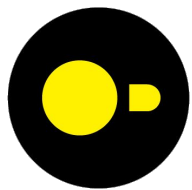


Moondog

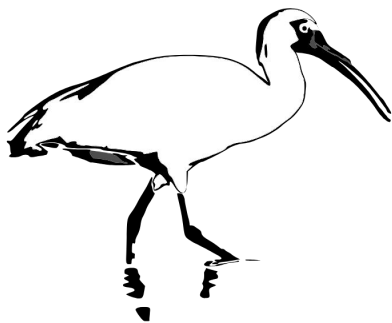


Michael Chow

How do you work with data in Python?

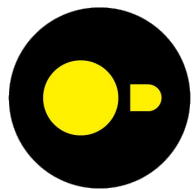


DuckDB



ibis

How do you work with data in Python?

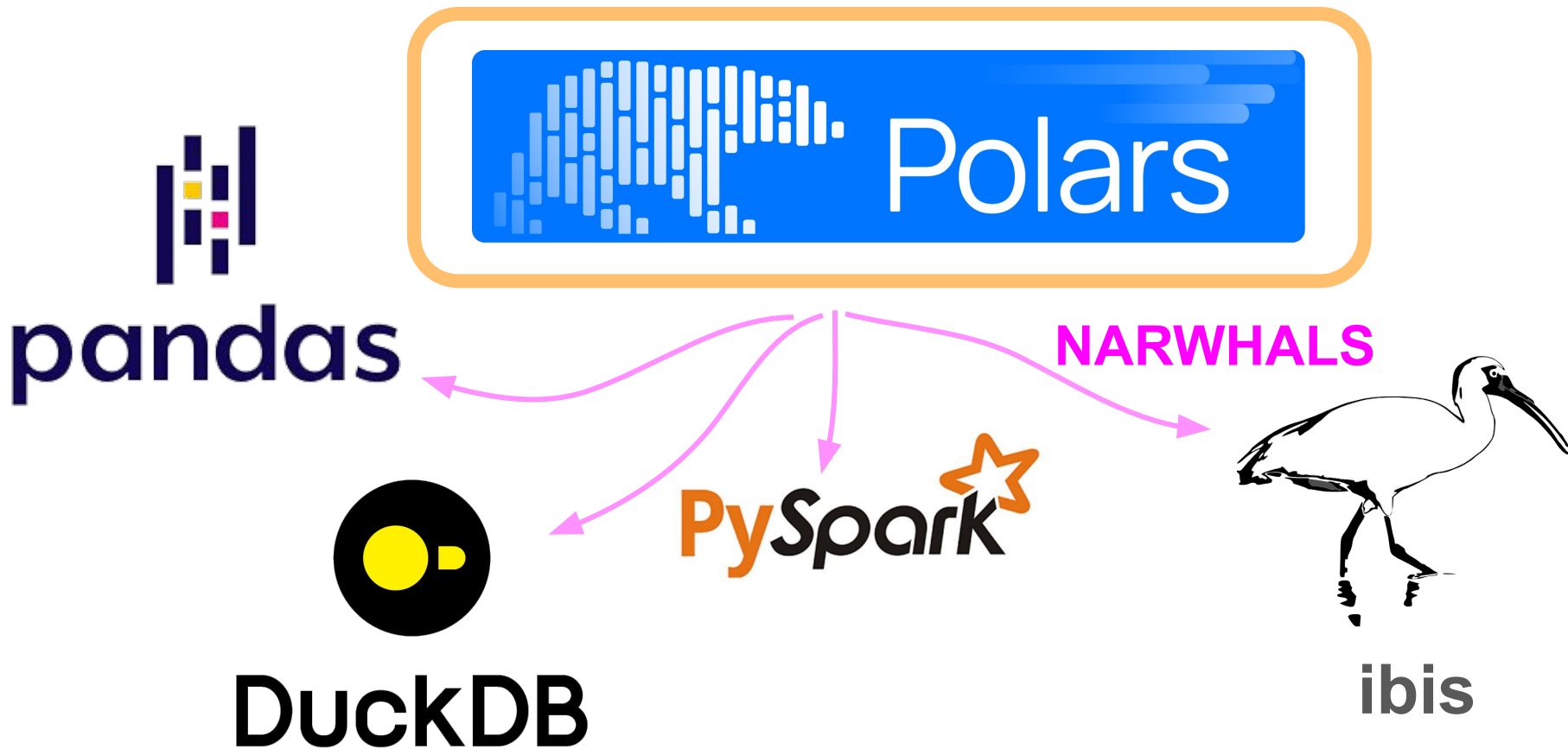


DuckDB



ibis

How do you work with data in Python?



Why Polars

- Polars: DataFrame library for pharmaverse-py
- Delightful: “Come for the speed, stay for the API”
- Everywhere: Can run on other DataFrames / databases

Polars: DataFrames for the Pharmaverse-py

What's the pharmaverse?

ADaM in R
(data and metadata)

PHARMAVERSE



gtsummary:
statistical tables

What's the pharmaverse?

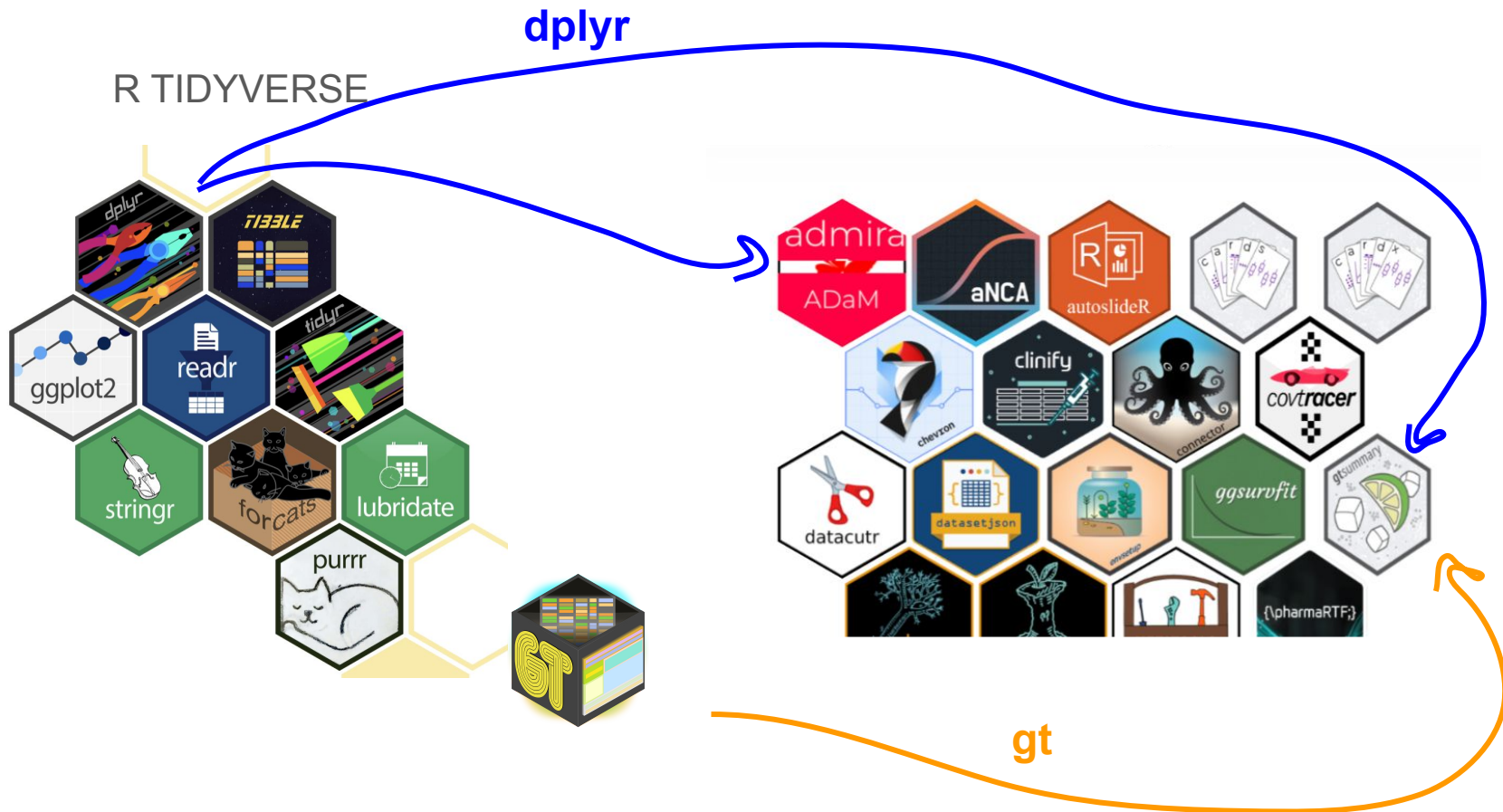
R TIDYVERSE



PHARMAVERSE



What's the pharmaverse?



What about Python?

Python



What about Python?

Python



Great Tables

narwhals



Pharmaverse workflow

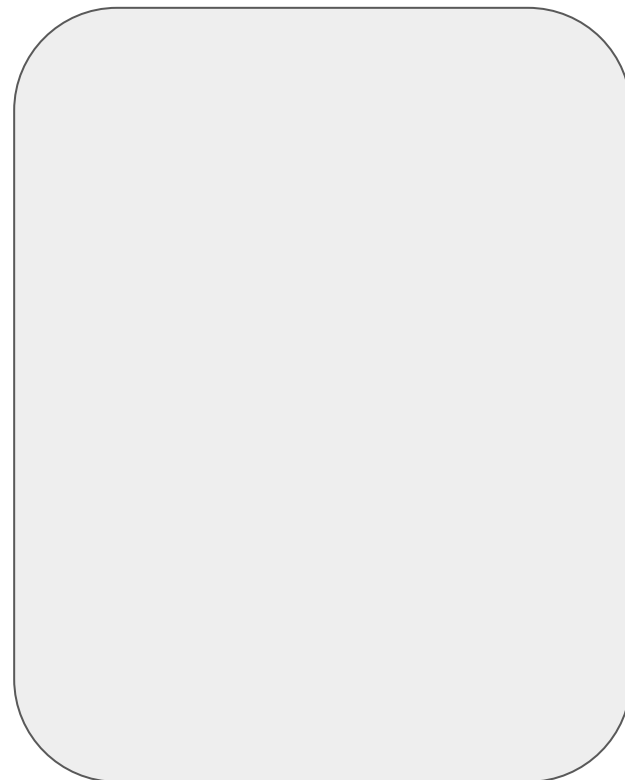
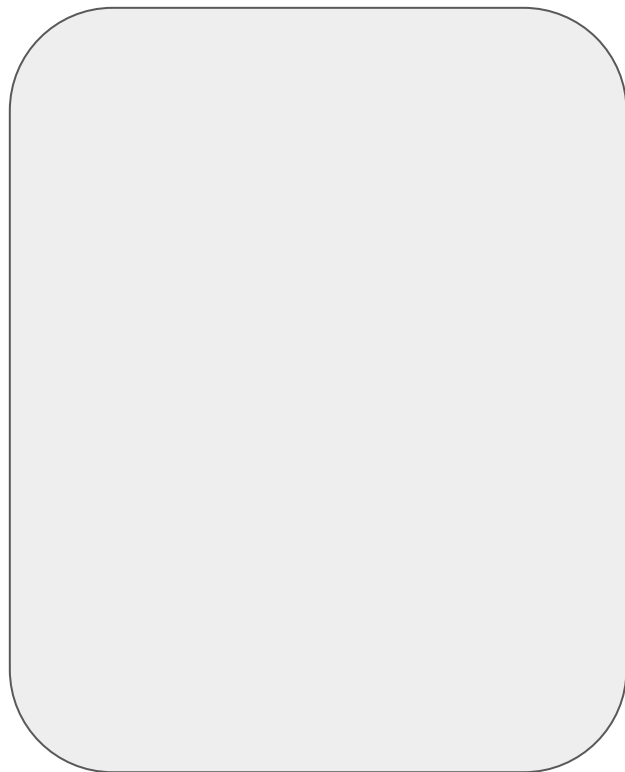
Python

R

Query database
(big data)

Wrangle data

TFLs



Pharmaverse workflow

Python

Query database
(big data)

Wrangle data

TFLs

R



gt / gtsummary

Pharmaverse workflow

Python

Query database
(big data)

narwhals

Wrangle data



TFLs

Great Tables

R



gt / gtsummary

If you like R, you'll love Polars

Polars is delightful

```
import polars as pl
import narwhals as nw
import duckdb
```

```
df = pl.DataFrame({
    "x": [1, 2, 3],
})
```

```
(
    df
    .with_columns(z = pl.col("x") - pl.col("x").mean())
)
```

data

analysis

Create column named z

Lazy expressions

Polars runs everywhere (via narwhals)

```
import polars as pl
import narwhals as nw
import duckdb

df = pl.DataFrame({
    "x": [1, 2, 3],
})

(
    df
    .with_columns(z = pl.col("x") - pl.col("x").mean())
)
```

Polars runs everywhere (via narwhals)

```
import polars as pl
import narwhals as nw
import duckdb

df = pl.DataFrame({
    "x": [1, 2, 3],
})

(
    nw.from_native(df)
    .with_columns(z = nw.col("x") - nw.col("x").mean())
)
```

Now can
replace with
any data

Pandas
pyspark
duckdb
snowflake

Special narwhals
function

Swap in nw.col()

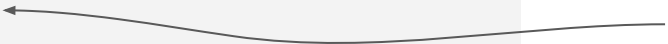
Polars is delightful (narwhals)

```
import polars as pl  
import narwhals as nw  
import duckdb
```

```
df = duckdb.sql("SELECT * FROM df")
```

```
(  
    nw.from_native(df)  
    .with_columns(z = nw.col("x") - nw.col("x").mean())  
)
```

duckdb



FAST

DELIGHTFUL

RUNS
EVERYWHERE

Polars is a delightful DataFrame library
that runs everywhere

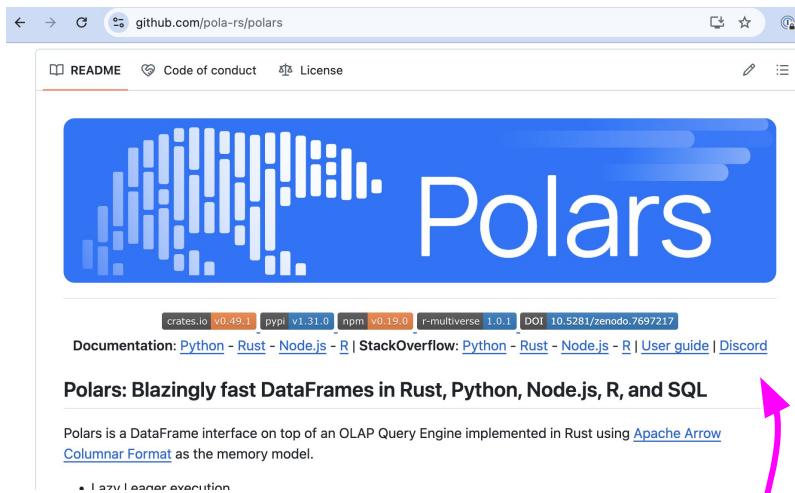
Thanks!



Sally Yang



Phil
Bowsher



discord