

Some

Tools, Libraries and Analyses in Biomedical Data Science

Nick Giangreco

PhD Candidate in Systems Biology at Columbia University

12/20/17

New York Healthcare Artificial Intelligence Society

<http://nickg.bio>

[GitHub repository for talk](#)

*Everything in red is a link

What's this about

- Disclaimer: NOT COMPREHENSIVE, MY VIEWS NOT Columbia's, TRAINING IN-PROGRESS

What's this about

- Disclaimer: NOT COMPREHENSIVE, MY VIEWS NOT Columbia's, TRAINING IN-PROGRESS
- Tools, analyses and libraries I use

What's this about

- Disclaimer: NOT COMPREHENSIVE, MY VIEWS NOT Columbia's, TRAINING IN-PROGRESS
- Tools, analyses and libraries I use
- **GitHub repository**
 - Code notebooks
 - Observational Health data

What you'll get out of this talk

- Links to resources

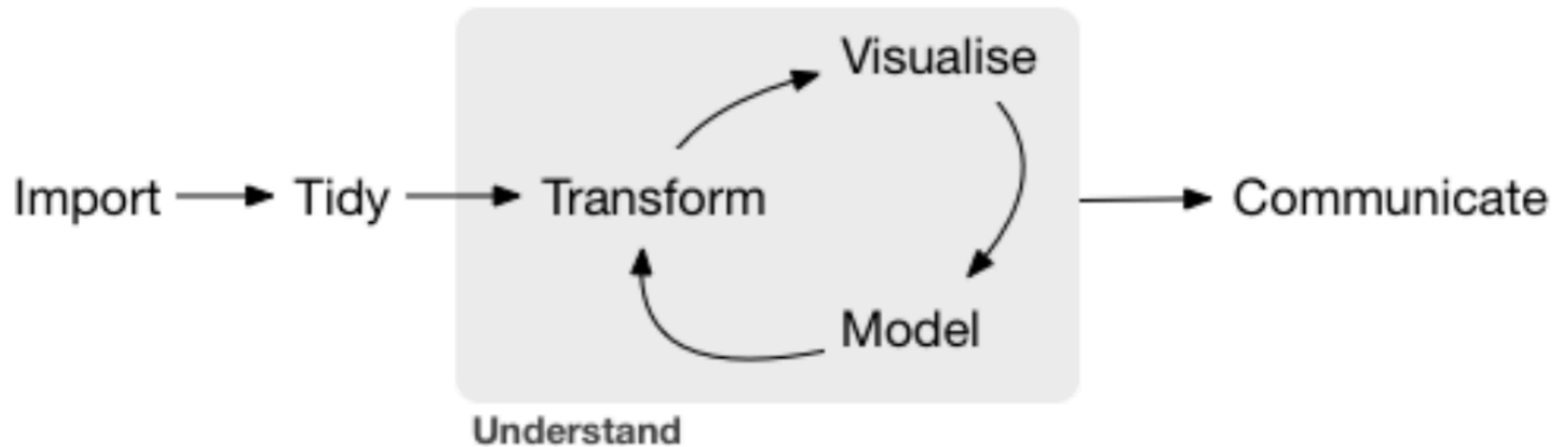
What you'll get out of this talk

- Links to resources
- Use-case example
 - Code
 - Notebooks

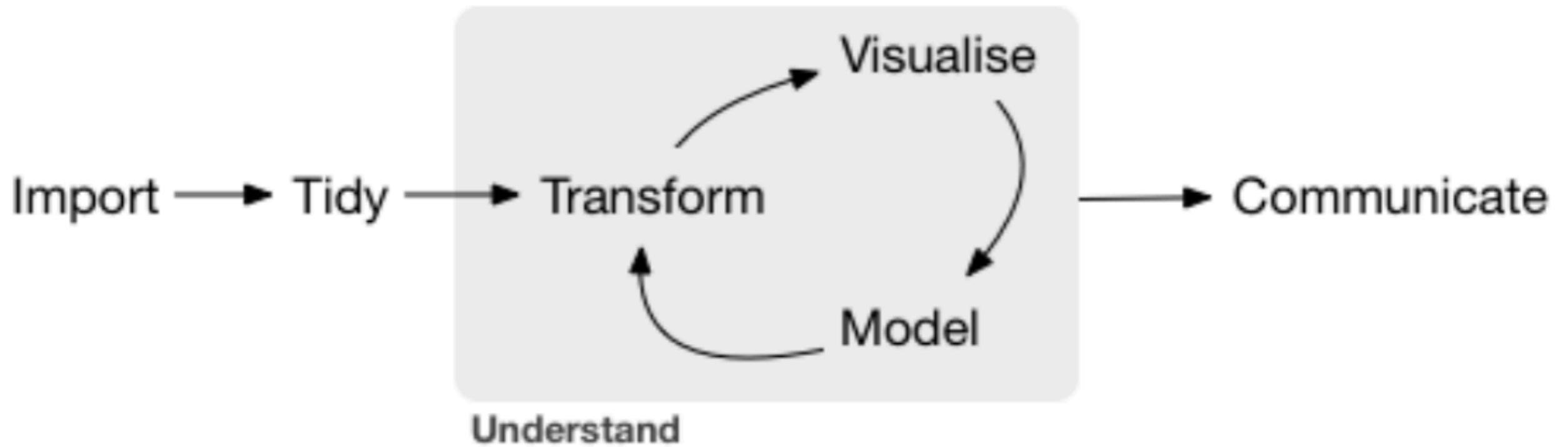
What you'll get out of this talk

- Links to resources
- See use-case example
 - Code
 - Notebooks
- “Published” [GitHub repository](#)
 - open, observational health data
 - Sample notebooks

Data Science Workflow



Tools:



Tools:



Libraries:

A Common Database Interface (DBI)

DBI package - [CRAN.R-project.org](https://cran.r-project.org/web/packages/DBI/README.html)

<https://cran.r-project.org/web/packages/DBI/README.html>

The **DBI** package defines a common interface between the **R** and database management systems (DBMS). The interface defines a small set of classes and methods similar in spirit to Perl's **DBI**, Java's **JDBC**, Python's **DB-API**, and Microsoft's **ODBC**. It defines a set of classes and methods defines what operations are ...

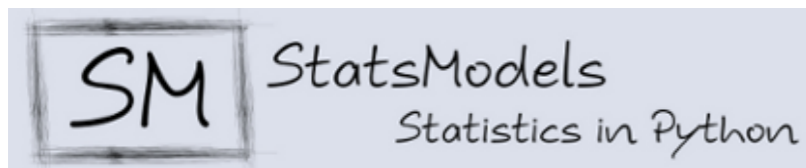
SciPy.org

Install Getting Started Documentation Report Bugs Blogs

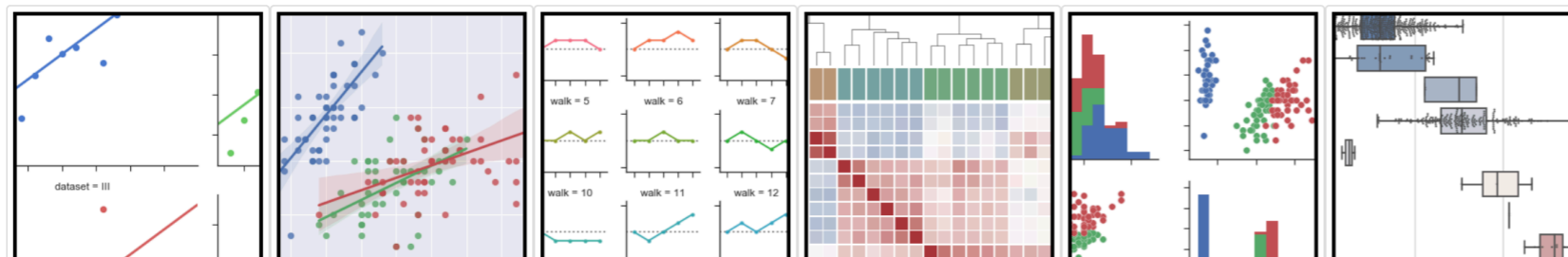
SciPy (pronounced "Sigh Pie") is a Python-based ecosystem of open-source software for mathematics, science, and engineering. In particular, these are some of the core packages:

- NumPy: Base N-dimensional array package
- SciPy library: Fundamental library for scientific computing
- Matplotlib: Comprehensive 2D Plotting
- IPython: Enhanced Interactive Console
- Sympy: Symbolic mathematics
- pandas: Data structures & analysis

More information...



seaborn: statistical data visualization



Use-case examples

Sorry, I'm skipping a lot of background...
Click [this](#), and see the presentation slides 33-50.
Also see this [GitHub repository](#)

ADVERSE DRUG REACTIONS

- ▶ Drugs are supposed to treat your disease
- ▶ But, not all drugs work the same
- ▶ May cause an **adverse reaction**,
 - ▶ fever, nausea, bloating, dizziness/confusion
 - ▶ heart attack, liver damage, kidney failure, seizures

DATA

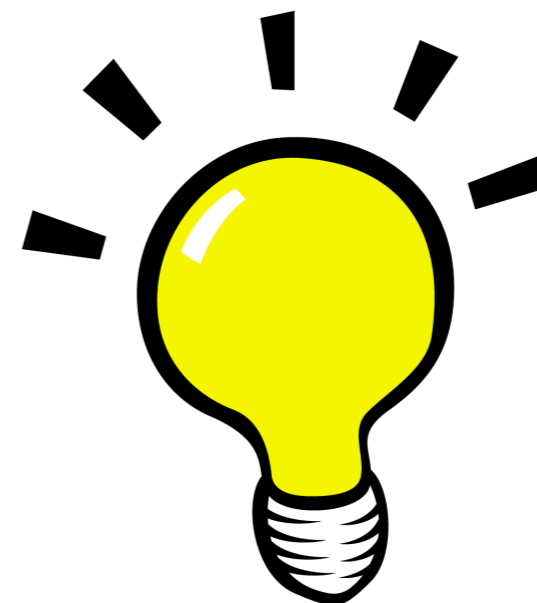
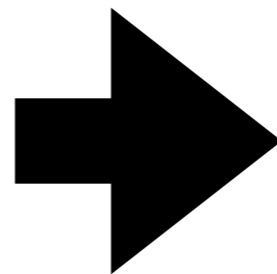


Public reports of adverse drug reactions

DATA

- ▶ Reports submitted indicate:
 - ▶ Drug(s) a patient was taking
 - ▶ Adverse reactions the patient experienced
 - ▶ Age of the patient

- ▶ Objective:



>8 million reports from 2004-2015

Computational Notebooks

- R markdown

Computational Notebooks

- R markdown
- Jupyter Notebooks

Demos

Thanks!