

DataOps

Data Science Empowerment through
DevOps, Cloud Computing and Building
your own Applications

Kelly O'Briant

Data Science Product Engineer

kelly@rladies.org

@kellrstats | @RLadiesDC



- Business Science University course developer



- R-Ladies Washington DC Chapter Founder and Organizer
- R-Ladies Global unofficial “cloud expert”



- Publish a monthly series called `.rprofile` on the rOpenSci blog

My Talk Goal:

I want you to leave this conference so excited, you go back to work and completely ignore whatever project you're supposed to be working on because you're so pumped up about building a data product and you can't stop yourself from doing it.

Motivation

Why I talk about Data Science Empowerment

R-Ladies events

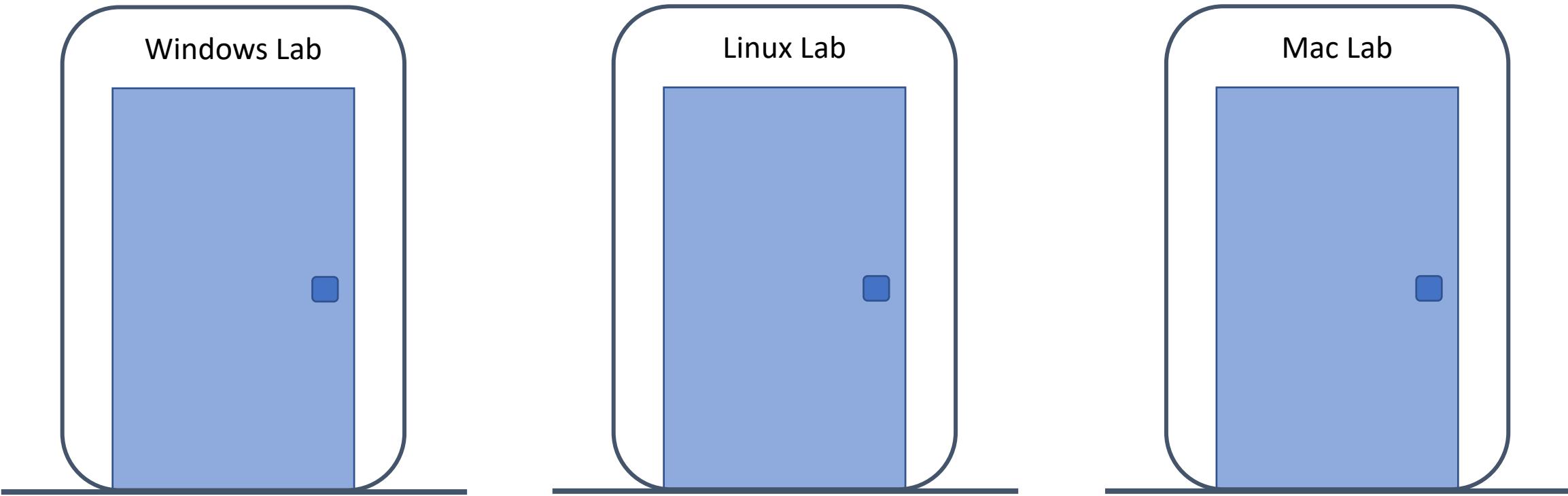
- How do I get a job as a data scientist/analyst/anything?
- What should I study/learn/do/produce to be a data scientist?
- Am I even a data scientist? Is what I do data science?

Why are data products empowering?

- I use data products to justify/prove to myself that I belong, that my ideas are valid and to help me communicate with people who are bad at listening (or when I'm bad at speaking)

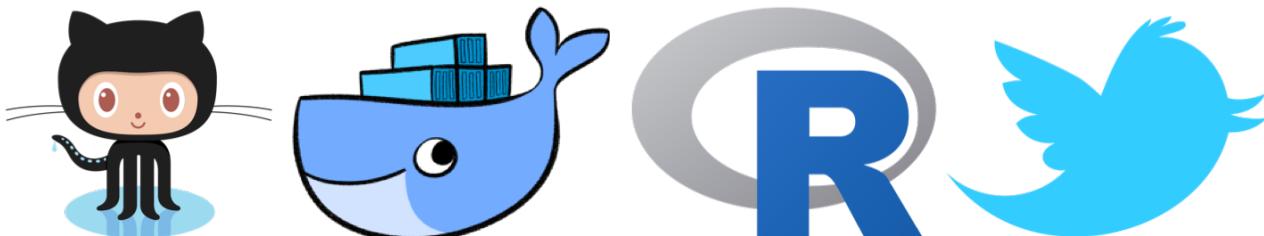
Motivation

Traumatic Experiences!



R-Ladies + International Women's Day Twitter Campaign

- Create a twitter bot using **R code** to tweet out a profile for every woman in our Global speaker directory
- Project collaboration through **GitHub**
- **Docker** linked to a local volume
- **Twitter Application(s)**



IWD 2018 featured R-Ladies
@rladies_iwd2018 Following

Read more about how amazing this R-Lady is! rladies.org/ladies-complet... #rladies #iwd2018 ♀

KELLY O'BRIANT MSC COMPUTATIONAL SCIENCES
Data Scientist
B23
City
Washington D.C.
United States

A portrait photo of a woman with long brown hair, smiling.

BIOGRAPHICAL INFO

- **Interests:** sparklyr, shiny, R-Ladies DC
- **Contact method:** Twitter

Categories: Speaker, United States

UPDATED 1 MONTH AGO.

2:55 AM - 8 Mar 2018

Deploy and Use H2O Machine Learning Models in Production

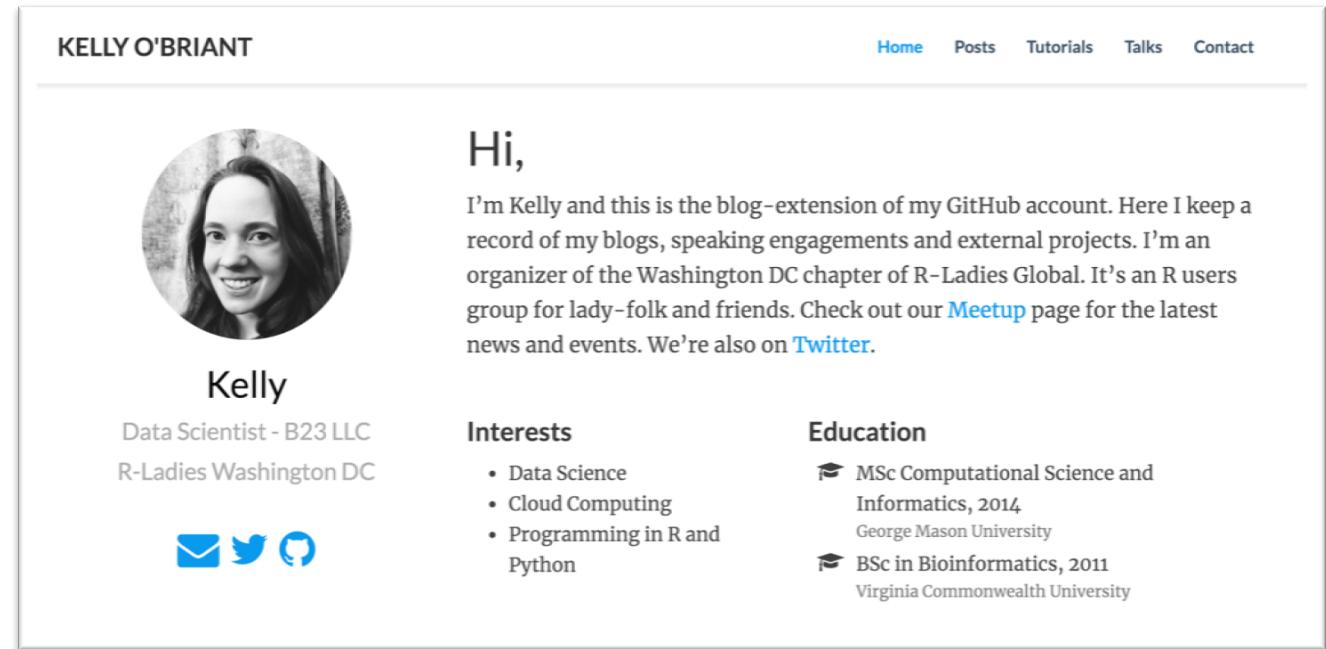
- Build and validate a model in python working in a **Jupyter Notebook** with the **H2O machine learning API**
- Package the model code as a POJO or MOJO file
- Deploy the model to **H2O.ai STEAM** to create an ML prediction service complete with a **REST API** query URL

A screenshot of a web browser showing the "Prediction Service" interface on localhost:55001. The interface is divided into sections: "MODEL INPUT PARAMETERS" (containing dropdown menus for Month, DayofMonth, DayOfWeek, DepTime, UniqueCarrier, Origin, Dest, and Distance), "Query String" (describing how input parameters are converted into a REST API query string), "PREDICTION RESULTS" (showing "Model Predictions" for Index 0 (Label N, Probability 0.6941) and Index 1 (Label Y, Probability 0.3059)), and "Model Runtime Stats" (showing the service started at 2016-07-15 22:52:45 UTC and uptime of 3 m 37 s).

Index	Labels	Probability
0	N	0.6941
1	Y	0.3059

Create and Maintain a Personal Website

- Use the **blogdown** package in an **RStudio project** to create the framework for a **Hugo** static website
- Create content for the site by writing **Rmarkdown** files
- Compile and deploy the static site – choose a hosting mechanism: **GitHub**? Continuous Integration with **Netlify**?

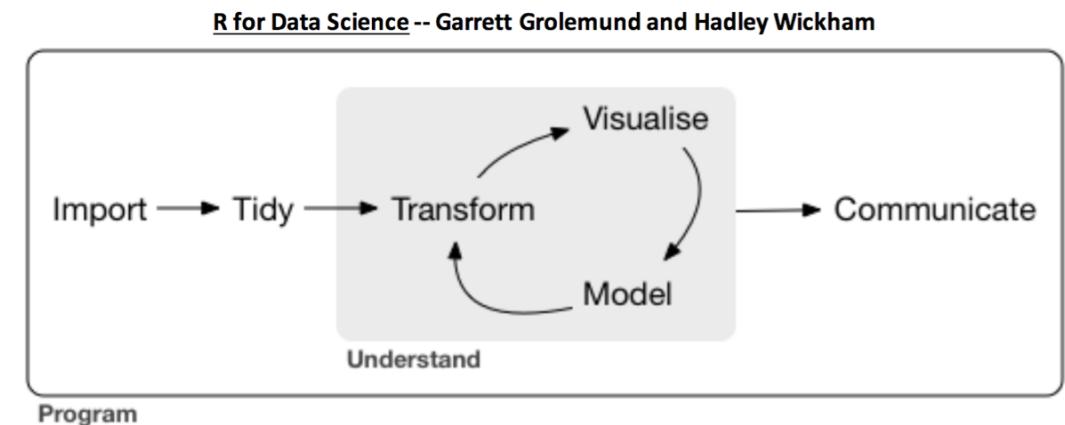
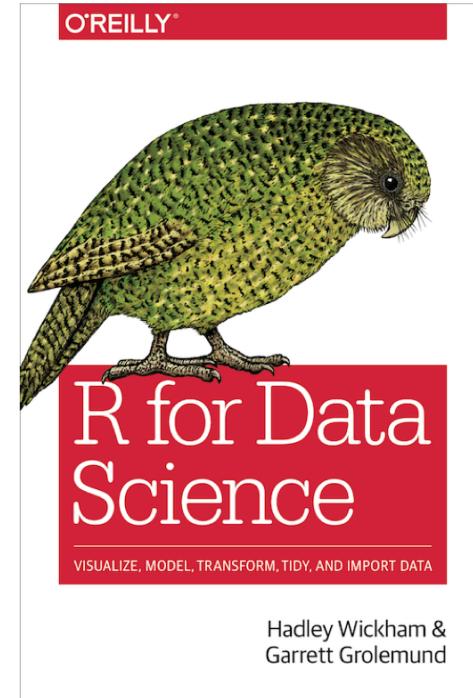


The screenshot shows a clean, modern website layout. At the top right is a navigation bar with links to Home, Posts, Tutorials, Talks, and Contact. The main content area features a large, circular profile picture of a woman with long dark hair, smiling. Below the photo is her name, "Kelly". Underneath her name are her titles: "Data Scientist - B23 LLC" and "R-Ladies Washington DC". To the right of her bio are three social media icons: an envelope for email, a Twitter bird, and a GitHub logo. To the right of the bio is a friendly greeting: "Hi," followed by a paragraph about the author, Kelly O'Briant. The paragraph describes her as a Data Scientist at B23 LLC and a member of R-Ladies Washington DC. It also mentions her role as an organizer for R-Ladies Global and provides links to their Meetup page and Twitter account. On the far right, there are two sections: "Interests" (listing Data Science, Cloud Computing, Programming in R and Python) and "Education" (listing her MSc in Computational Science and Informatics from George Mason University in 2014 and her BSc in Bioinformatics from Virginia Commonwealth University in 2011).



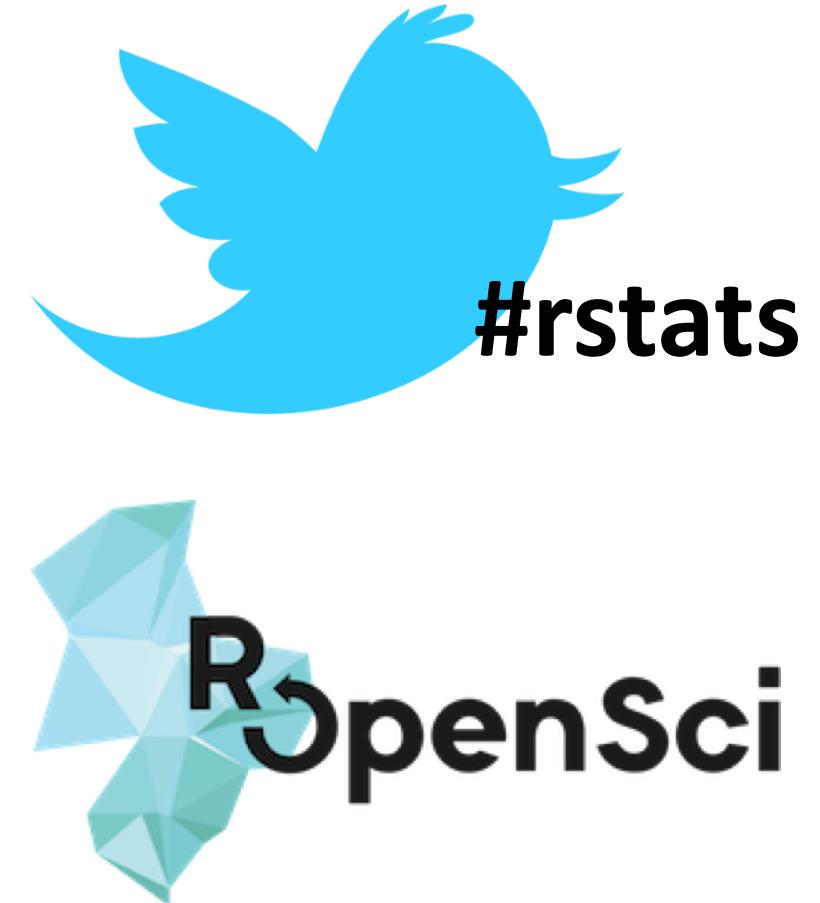
Why are you so into R?

- It's great for Data Science
- The community at large is awesome
- The female community is awesome
- R integrates with other tech
- It's growing really fast in cool ways
- I can use it to build cool stuff



Why are you so into R?

- It's great for Data Science
- The community at large is awesome
- The female community is awesome
- R integrates with other tech
- It's growing really fast in cool ways
- I can use it to build cool stuff



Why are you so into R?

- It's great for Data Science
- The community at large is awesome
- **The female community is awesome**
- R integrates with other tech
- It's growing really fast in cool ways
- I can use it to build cool stuff

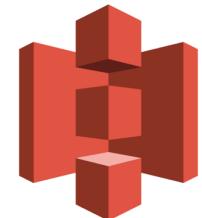
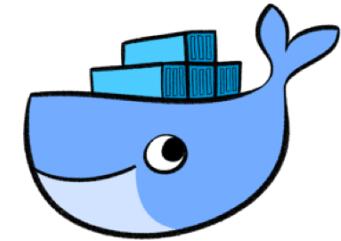


Worldwide organization
that **promotes gender diversity**
in the **R** community via **meetups**
and mentorship in a **friendly** and
safe environment



Why are you so into R?

- It's great for Data Science
- The community at large is awesome
- The female community is awesome
- R integrates with other tech
- It's growing really fast in cool ways
- I can use it to build cool stuff



Amazon S3



REST API



amazon
REDSHIFT

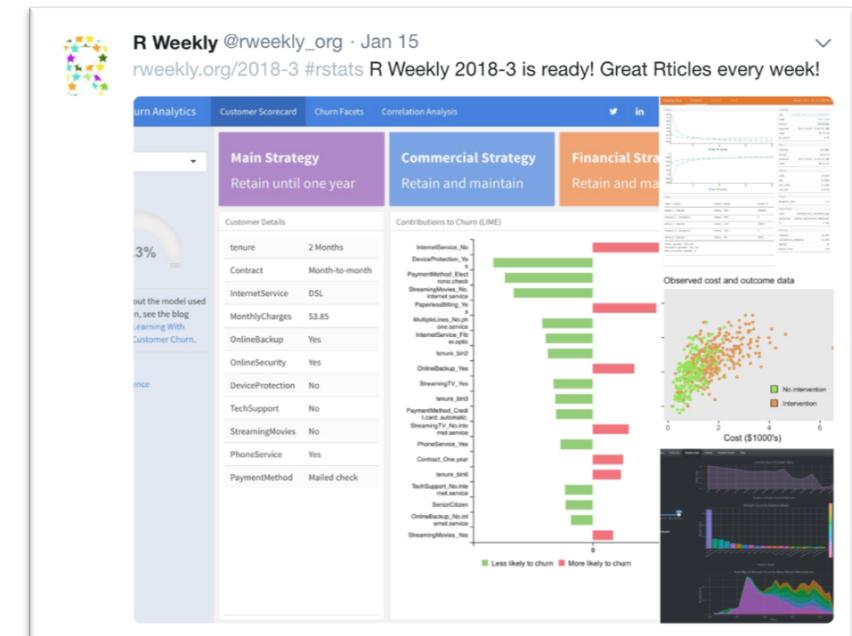
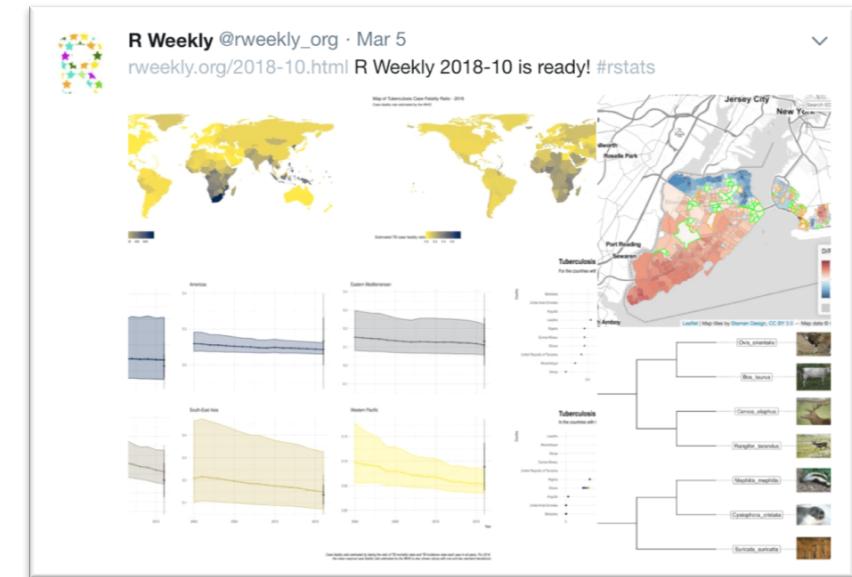


A Fast & Modern Static Website Engine



Why are you so into R?

- It's great for Data Science
- The community at large is awesome
- The female community is awesome
- R integrates with other tech
- **It's growing really fast in cool ways**
- I can use it to build cool stuff



Why are you so into R?

- It's great for Data Science
- The community at large is awesome
- The female community is awesome
- R integrates with other tech
- It's growing really fast in cool ways
- I can use it to build cool stuff



Back to the topic: DataOps

1. It usually takes a little DevOps to build a Data Product
2. Building more Data Products is empowering – good for your portfolio and soul

What is DevOps

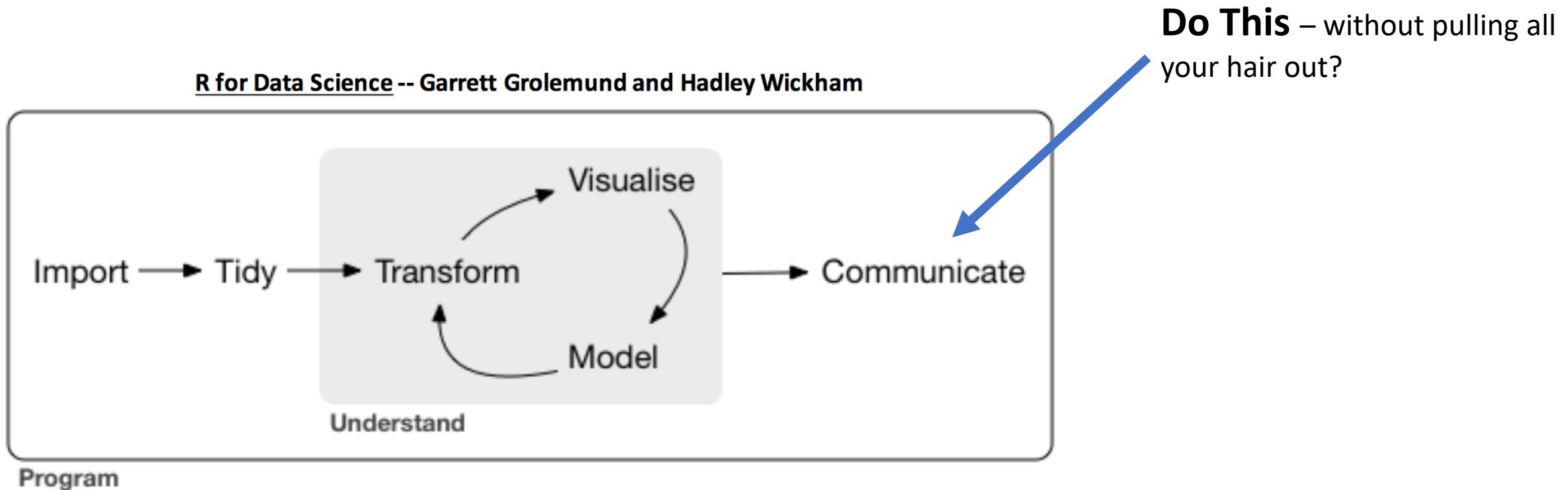
And why should Data-oriented people care about it?

DevOps is...

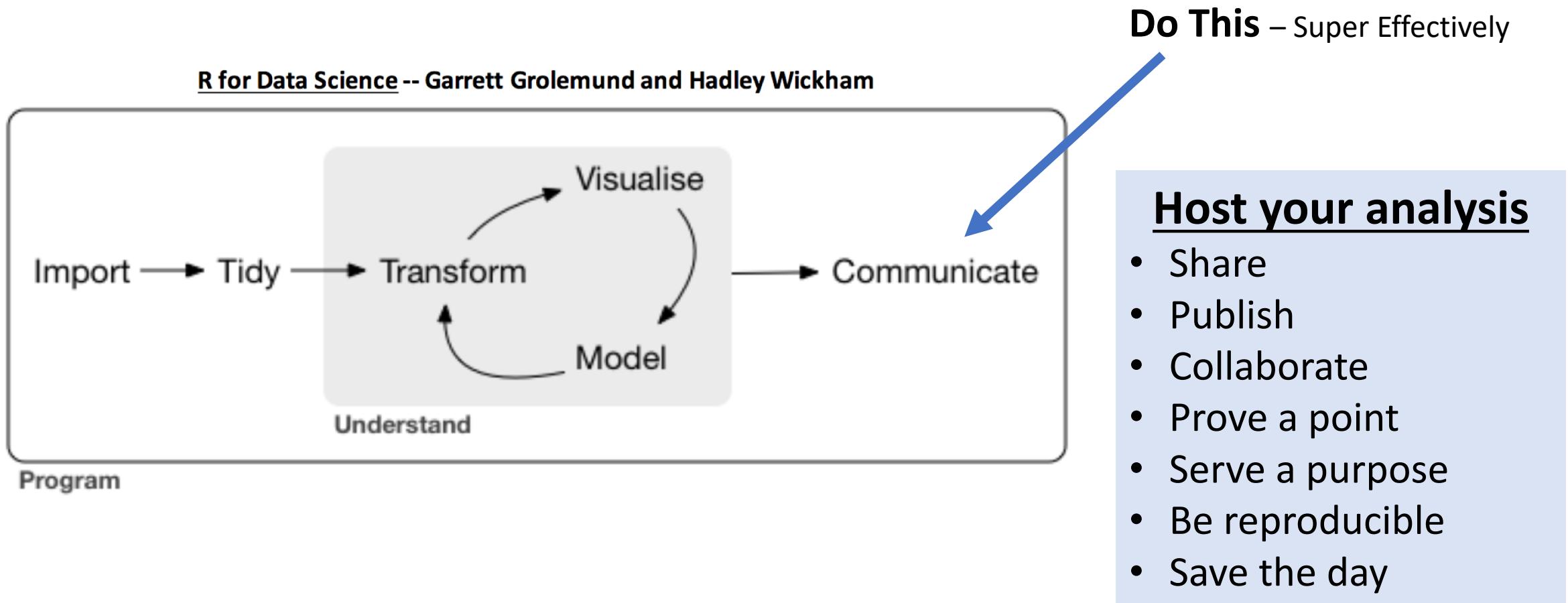
“A combination of cultural philosophies, practices and tools that increases an organizations ability to deliver applications and services at high velocity.

- AWS DevOps Blog

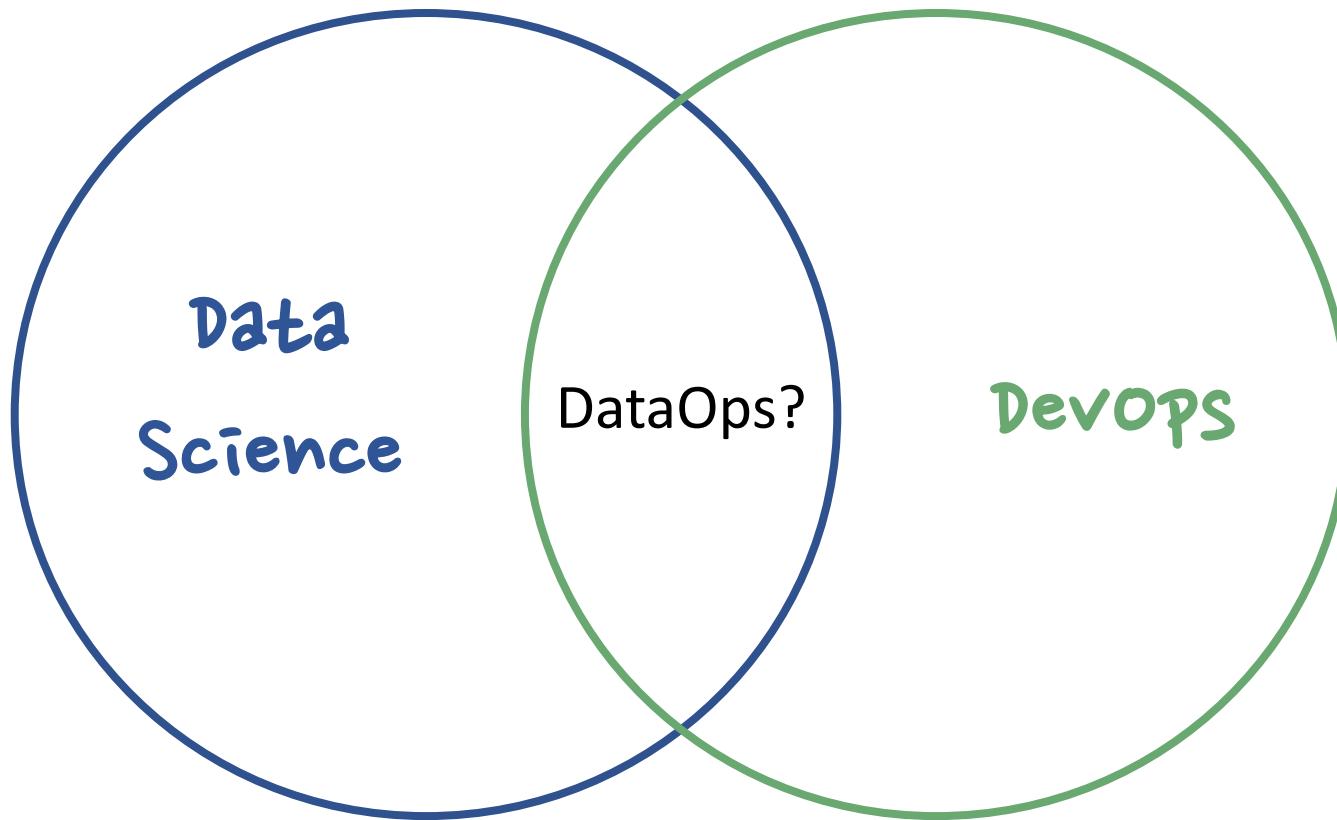
Deliver applications and services at high velocity



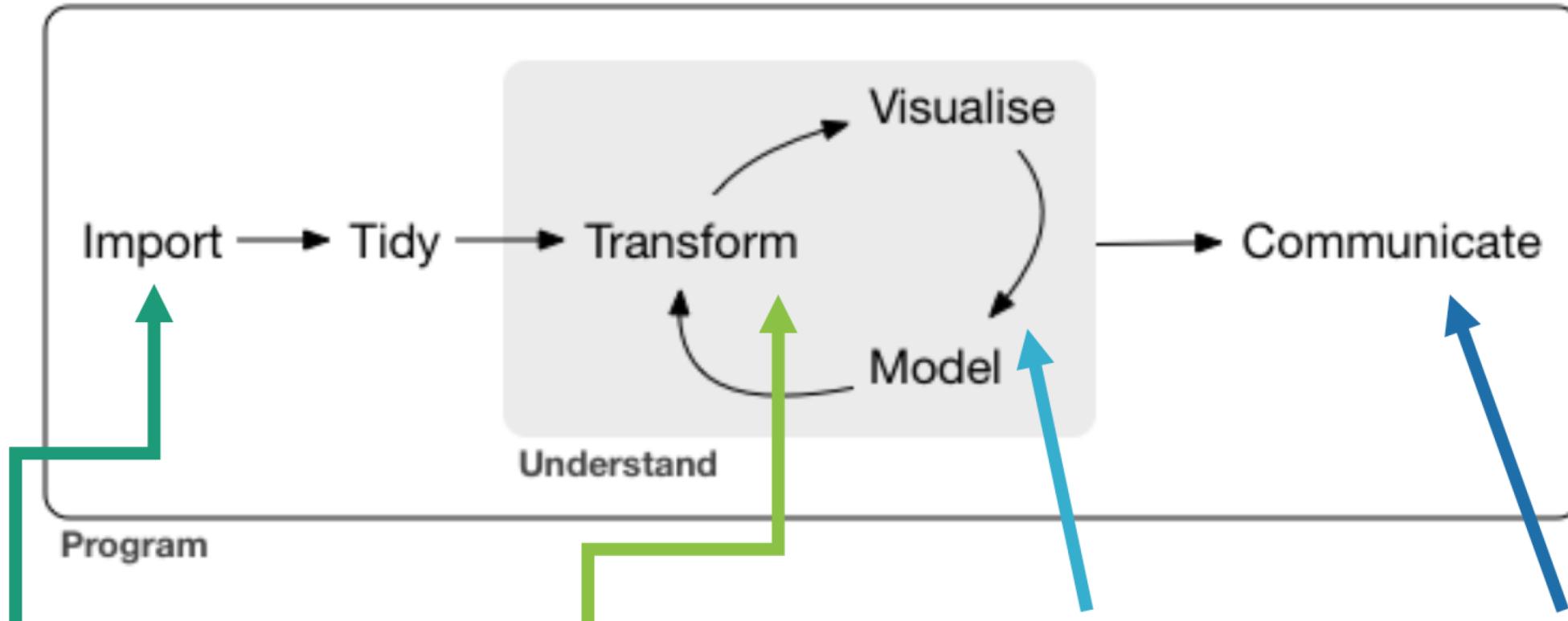
Deliver applications and services at high velocity



What is DataOps?



Anywhere you can put a little DevOps magic into your data science workflow



Get access to the data you need

- Multiple sources?
- Coordinated effort?

Discover you need more data
or you need programmatic access to data regularly

Have collaborators validate your models, methods or preliminary findings

Host your analysis
share, publish, prove a point, serve a purpose, be reproducible

Build More Data Products

So that you and others can use them to solve real problems

Try Shiny!

An R package for building interactive web applications



The Iris Dataset

Codebook for iris

Autogenerated data summary from dataMaid

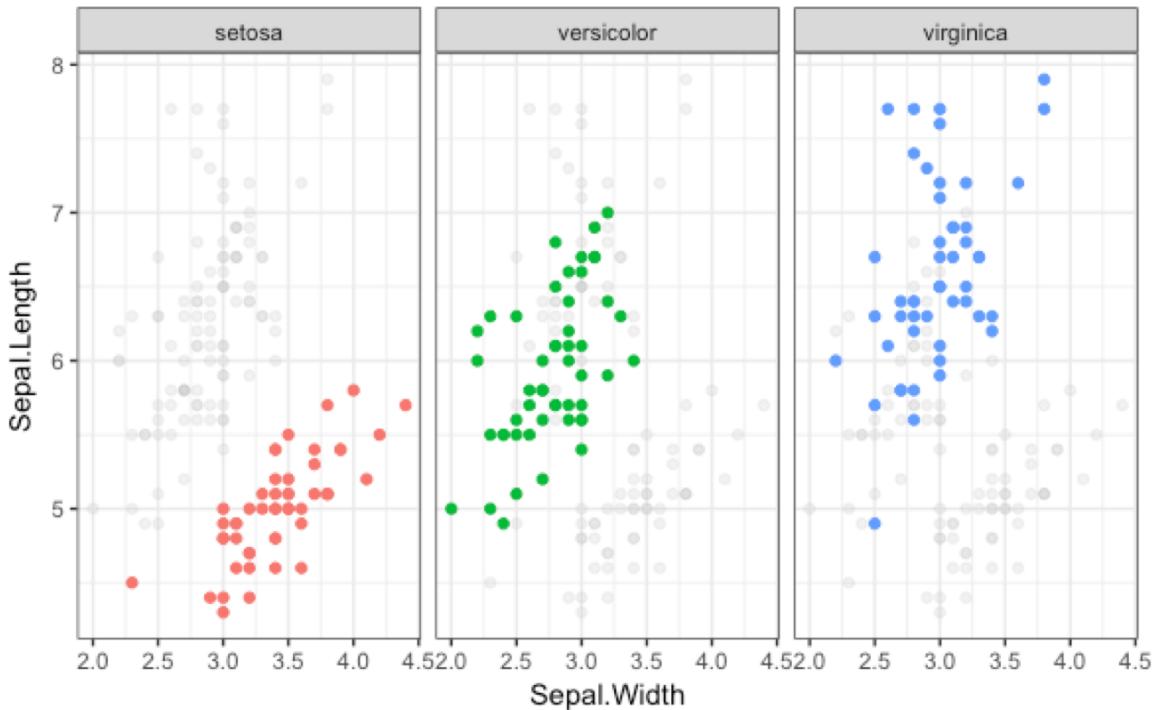
Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	150
Number of variables	5

Codebook summary table

Label	Variable	Class	# unique values	Missing
	Sepal.Length	numeric	35	0.00 %
	Sepal.Width	numeric	23	0.00 %
	Petal.Length	numeric	43	0.00 %
	Petal.Width	numeric	22	0.00 %
	Species	factor	3	0.00 %



Do Machine Learning!

So Hot Right Now



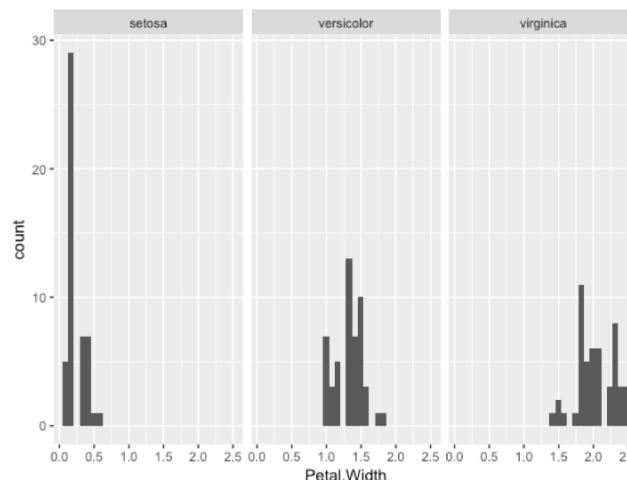
**What Species
is this iris??**

Credit: xkcd

1. Turn your ideas into R code

- Write functions to generate the plots you're envisioning
- Package: ggplot2
- Train and validate a machine learning model to use
- Package: caret

```
geom_hist_basic <- function(var){  
  ggplot(iris, aes_string(x = var)) +  
    geom_histogram() +  
    facet_wrap(~ Species)  
}
```



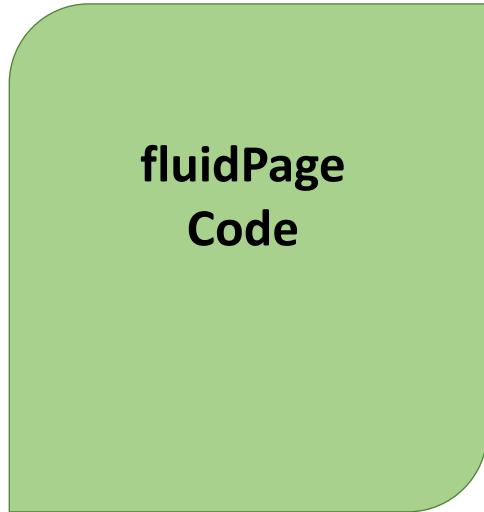
`predict_matrix(fit.knn, validation)`

Confusion Matrix and Statistics

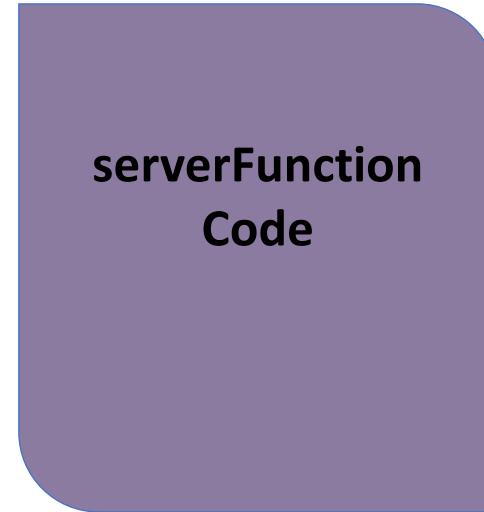
Prediction	setosa	versicolor	virginica
setosa	10	0	0
versicolor	0	8	1
virginica	0	2	9

2. Turn your R code into an R Shiny app

Client Side Code:
User Interface and
Input Elements



Server Side Code:
(Reactive) R Output
Elements



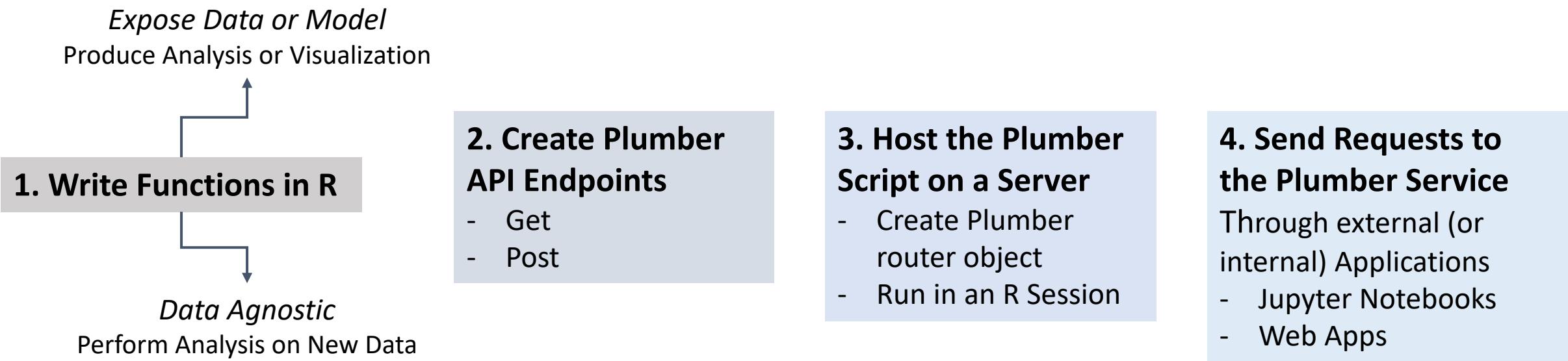
```
shinyApp(ui = fluidPage, server = serverFunction)
```

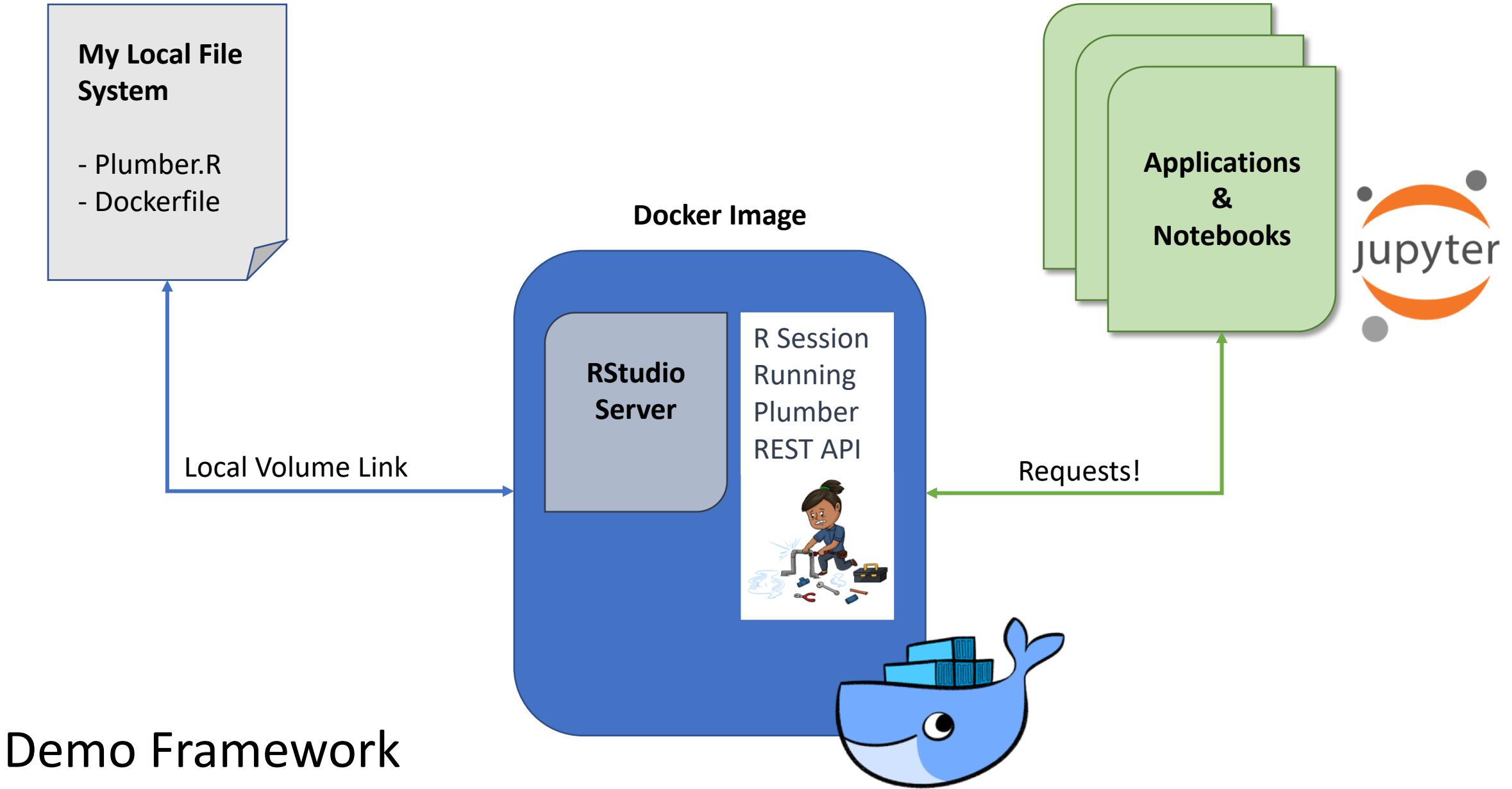
Try Plumber!

An R package for turning R functions into REST API endpoints



Let's Build a REST API with R





That's it!

Now go build some sweet data products

Resources for Learning R

If you're interested

R-Ladies Global Meetups

- Get involved!
- More female speakers, leaders, teachers, builders, friends!

R-Ladies ▾

Find out what's happening in R-Ladies Meetup groups around the world and start meeting up with the ones near you.

13,926 members | 66 Meetups

Related topics: [R Programming Language](#) · [Programming in R](#) · [R Project for Statistical Computing](#) · [Data Analytics](#) · [R Language](#) · [Data Science](#) · [Data Visualization](#) · [Statistical Computing](#) · [Women R Programmers](#) · [Applied Statistics](#)

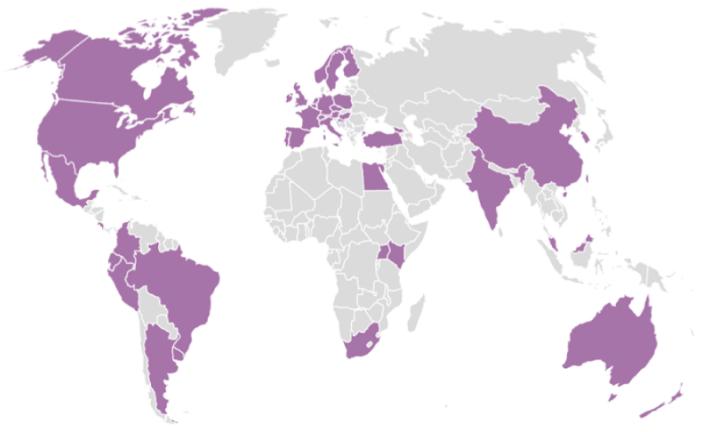


R-LADIES DIRECTORY

Looking for R-Ladies around the world? Check out our handy Directory!

You can browse our list of speakers for conferences & events, or view profiles of R-Ladies located by country!

[COMPLETE LIST](#)
[LIST OF SPEAKERS](#)
[↓ BY COUNTRY ↓](#)



RLadies.org
@RLadiesGlobal

RStudio Webinars

- All of the talks from RStudio::conf 2018 have just been published
- Highly recommend!

The screenshot shows the RStudio website's main navigation bar at the top, featuring links for Products, Resources (which is currently selected), Pricing, About Us, and Blogs, along with a search icon. Below the navigation, there are several sections: 'Learning Roadmap' with a list of topics like RStudio Essentials, RStudio Connect, Shiny Essentials, Advanced Shiny, The Essentials of Data Science, Advanced Data Science, and sparklyr; 'RStudio Events' listing rstudio::conf 2017 and rstudio::conf 2018; and a 'Latest Video' section featuring a thumbnail of J.J. Allaire speaking at a podium, titled 'Machine Learning with TensorFlow and R'. A 'VIEW MATERIALS' button is visible below the video thumbnail. At the bottom, a 'Learning Roadmap' section states: 'We've grouped our existing webinars into learning tracks that can help you expand your Data Science skills.'

R Studio®

Learning Roadmap

- RStudio Essentials
- RStudio Connect
- Shiny Essentials
- Advanced Shiny
- The Essentials of Data Science
- Advanced Data Science
- sparklyr

RStudio Events

- rstudio::conf 2017
- rstudio::conf 2018

Latest Video

Machine Learning with TensorFlow and R
J.J. Allaire

In this talk we'll explore the use of TensorFlow from R describing the complete workflow including data ingestion, training, and deploying models into production.

VIEW MATERIALS

We've grouped our existing webinars into learning tracks that can help you expand your Data Science skills.

Resources for Learning Shiny Development

shiny.rstudio.com

Learn Shiny

The tutorials on this page are primarily designed for users who are new to Shiny and want a guided introduction. If you use Shiny on a regular basis, you may want to skip these tutorials and visit the [articles](#) section where we cover individual Shiny topics at a more advanced level.

We created a [free course](#) on DataCamp to help you get started with Shiny! For an interactive, hands-on introduction to Shiny, make sure to check this course out.

The screenshot shows the DataCamp course landing page for "Building Web Applications in R with Shiny". It features the DataCamp logo and a "FREE COURSE" badge. The main title is prominently displayed. A yellow button labeled "Start Course For Free" is visible. To the right, there's a preview image of the course interface, which includes a hexagonal icon with the course title and a welcome message from a female instructor.

In addition, we also have a wealth of [videos](#) and [written tutorials](#).

Resources for Learning Plumber

www.rplumber.io

@TrestleJeff
on Twitter!



plumber Docs Examples ▾

 See code on GitHub

plumber

An R package that converts your existing R code to a web API using a handful of special one-line comments.



I'm Kelly! →
@kellrstats on Twitter



Note to self: Remember to give
out stickers

I have R-Ladies and R-Ladies Plumber Stickers!