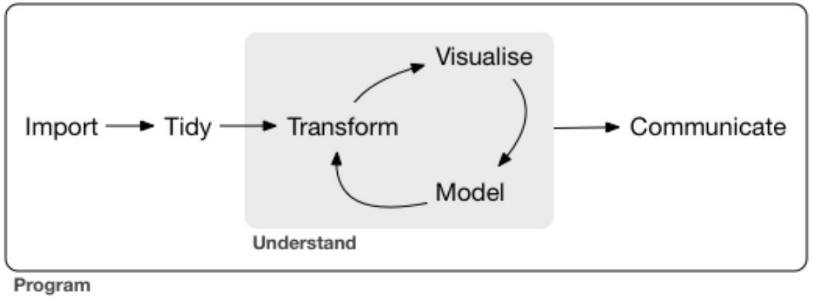
R in Production

Reflections on two years of solutions engineering at RStudio

What is Solutions Engineering?

RStudio

Industrial Research
Business Management
Human Resources
Government Work
Regulated Environments
Big Data Applications
Cloud Infrastructure
R in Production



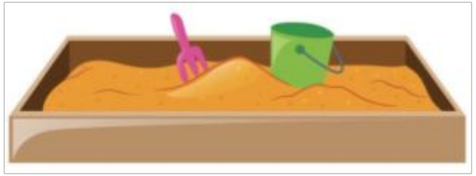
Solutions Engineers!

What is there to **learn**?
What are the **needs**?
What are the **problems**?

Stages of R in Production

Legitimacy →

Get R recognized as an analytic standard



Build a Sandbox
(Proof of Concept)

Competencies →

Understand and manage R tooling

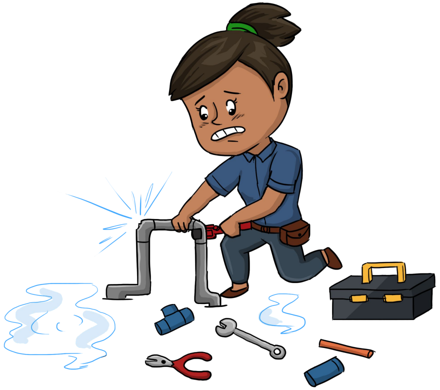
Invest in Learning
Develop Best Practices

Adoption

Rely on integrated R based solutions

Extend your Domain
Integrate and Interoperate

“R Admin” - Analytic Administrator Role



A data scientist who:

Onboards new tools, deploys solutions, supports existing standards

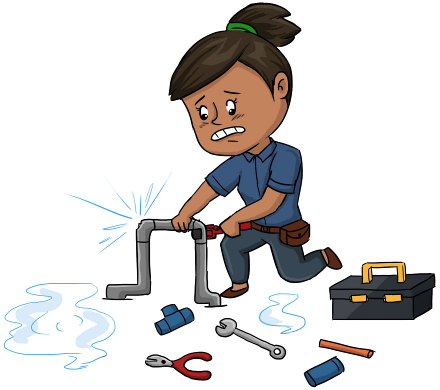
Works closely with IT to maintain, upgrade and scale analytic environments

Influences others in the organization to be more effective

Passionate about making R a legitimate analytic standard within the organization

Check out the RViews Blog:
Analytics Administration for R by Nathan Stephens

Times when I've felt entirely alone...



A data scientist who:

Onboards new tools, deploys solutions, supports existing standards

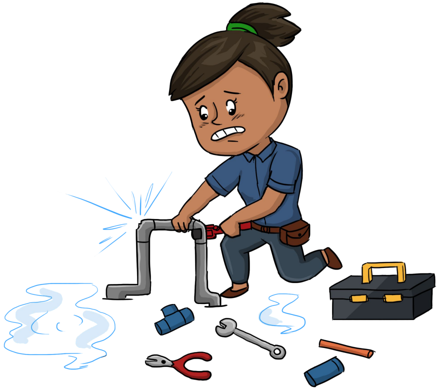
Works closely with IT to maintain, upgrade and scale analytic environments

Influences others in the organization to be more effective

Passionate about making R a legitimate analytic standard within the organization

Works at a giant tech company and Is very concerned with the improvement of daily work and analytic infrastructure

Times when I've felt entirely alone...



A data scientist who:

Onboards new tools, deploys solutions, supports existing standards

Works closely with IT to maintain, upgrade and scale analytic environments

Influences others in the organization to be more effective

Passionate about making R a legitimate analytic standard within the organization

Works at a tiny startup and Is very concerned with the improvement of daily work and analytic infrastructure

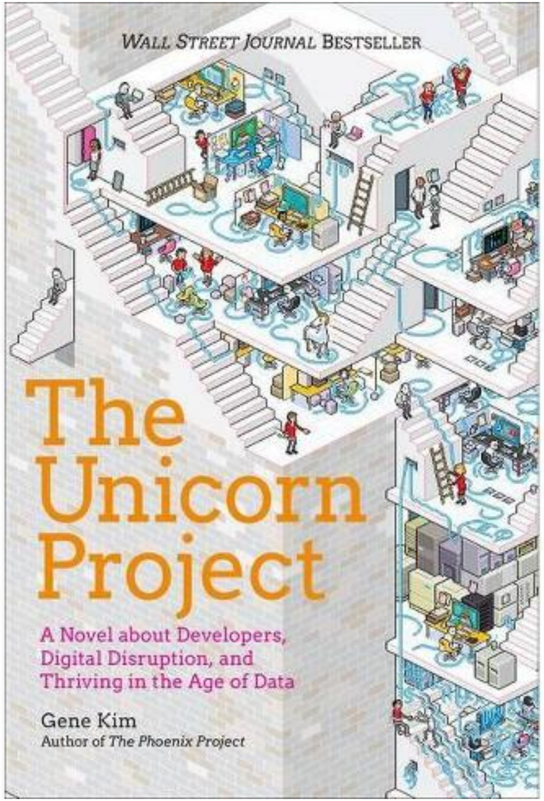
RStudio Solutions Engineering

is where I found my community

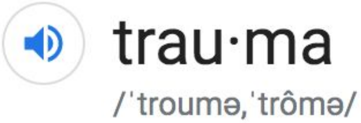
Phoenix is the most important project in the company. They've spent $20M over three years. And yet, here she is, trying to help, and they won't spend $5k on more disk space. And now she won't get a Dev environment for five months!

She buries her head in her hands and silently screams down at her keyboard.

...None of the meetings on her calendar seem interesting anymore. It's just people complaining about waiting. Waiting for something. Waiting for someone. Everyone is just waiting. And she wants no part of it right now.


WALL STREET JOURNAL BESTSELLER
The Unicorn Project
A Novel about Developers, Digital Disruption, and Thriving in the Age of Data
Gene Kim
Author of The Phoenix Project

Trauma


trau·ma
/ˈtroumə,ˈtrômə/

noun

1. an interpretation of an experience tied to a severely painful emotion

You need a team of **empathetic witnesses**. You need people to encourage you to keep going - to encourage your work when others don’t understand.
- Benjamin Hardy, PhD

The Five Ideals from The Unicorn Project by Gene Kim

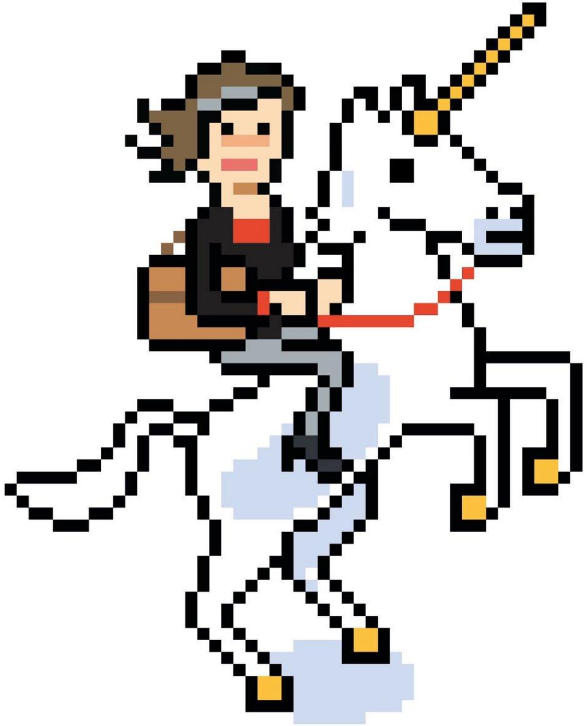
1 - Locality & Simplicity

2 - Focus, Flow & Joy

3 - Improvement of Daily Work

4 - Psychological Safety

5 - Customer Focus



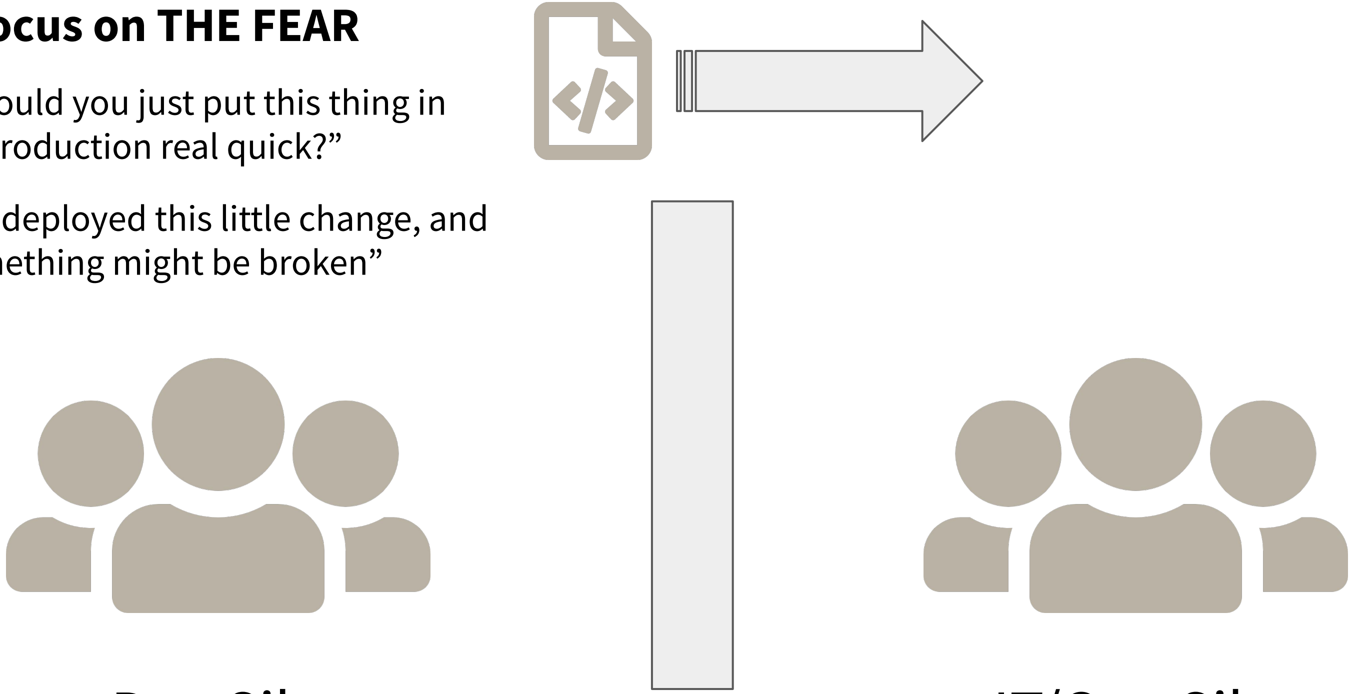
Ideal 1 - Locality & Simplicity

Classic DevOps Silo Diagram

Focus on THE FEAR

“Hey - could you just put this thing in production real quick?”

“Uh.. I just deployed this little change, and something might be broken”



Dev Silo

IT/Ops Silo

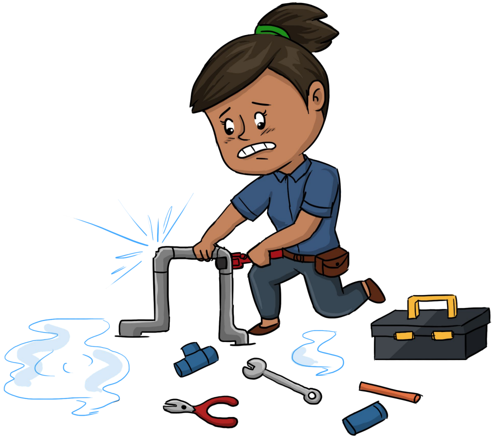
Can Data Scientists…

Independently develop, test, and deploy value to customers?

- Should data scientists be trusted with this responsibility?
- Who are the customers in this situation?
- What does deploying value entail?

The First Ideal: To what degree do teams have the capabilities and the authority to get what they need done? - Gene Kim

Challenges for the R User



Organizational

- Legitimizing R
- Working with IT

Technical

- Experience
- Education
- Exposure

Credibility Crisis Management Plan

Independently develop, test, and deploy value to customers

Local Development

EDA, Prototyping, Iteration

The “Hour-Long-Talk” of Data Products

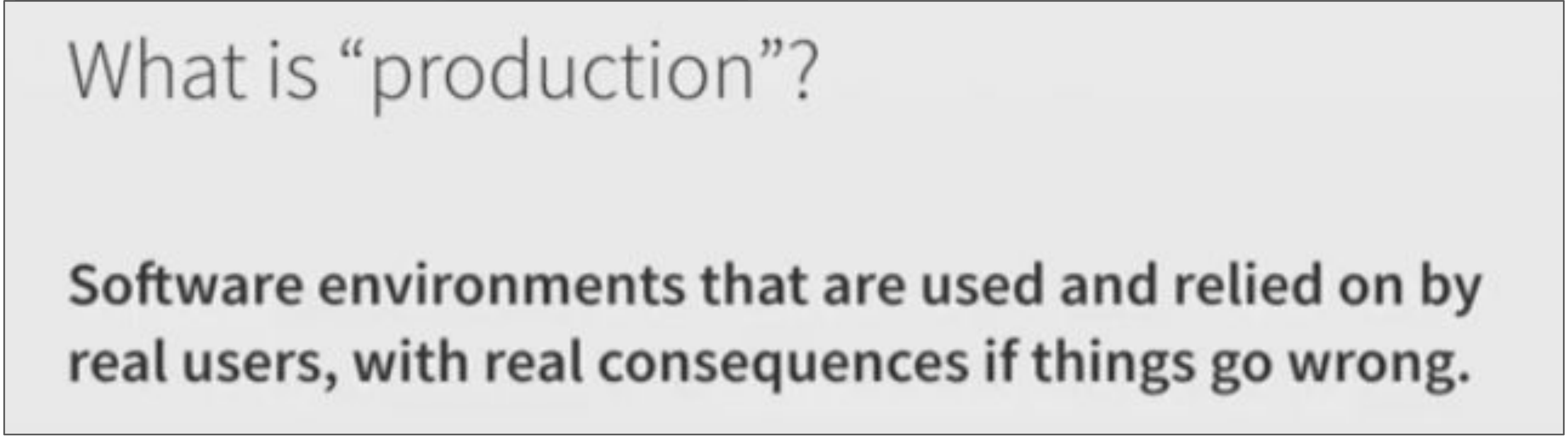
- Rambling, Cluttered
- Parts that work well
- Parts that work not-so well

Production Development

The “Lightning-Talk” of Data Products

- Targeted
- Elegant
- Streamlined
- Optimized

Joe Cheng “Shiny in Production”
RStudio Conf 2019 Keynote



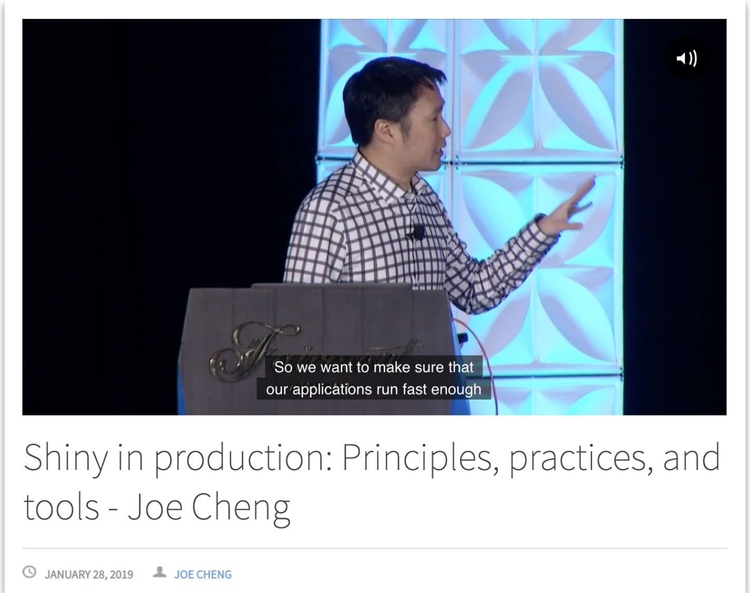
What is “production”?

Software environments that are used and relied on by real users, with real consequences if things go wrong.

Turn a Prototype into a Production Application

Performance Workflow

1. Use **shinyloadtest** to see if app is fast enough
2. If not, use **profvis** to see what’s making it slow
3. Optimize
   a. Move work out of shiny (very often)
   b. Make code faster (very often)
   c. Use caching (sometimes)
   d. Use async (occasionally)
4. Repeat!



So we want to make sure that
our applications run fast enough

Shiny in production: Principles, practices, and tools - Joe Cheng

JANUARY 28, 2019 JOE CHENG

Developing Trust is Challenging

Start by answering some questions…

- What is a Shiny Application?
- Who is the audience?
- What is your service level agreement definition? (SLA)
- What does your analytic architecture look like today?
  - What are your goals for evolving this architecture?
- How will monitoring be handled?
- Who is responsible for maintenance?

What does ‘Production’ mean?

Keep it up: unplanned outages are rare or nonexistent

Keep it safe: data, functionality, and code are all kept safe from unauthorized users

Keep it correct: works as intended, provides the right answers

Keep it snappy: fast response times, ability to predict needed capacity for expected traffic

Make work visible, Define shared goals, Build a checklist, Iterate

R (& Shiny) in Production Journey

Code Profiling

Version Control

Testing

Deployment/Release

Access/Security

Performance Tuning

Shared Goal: Testing!

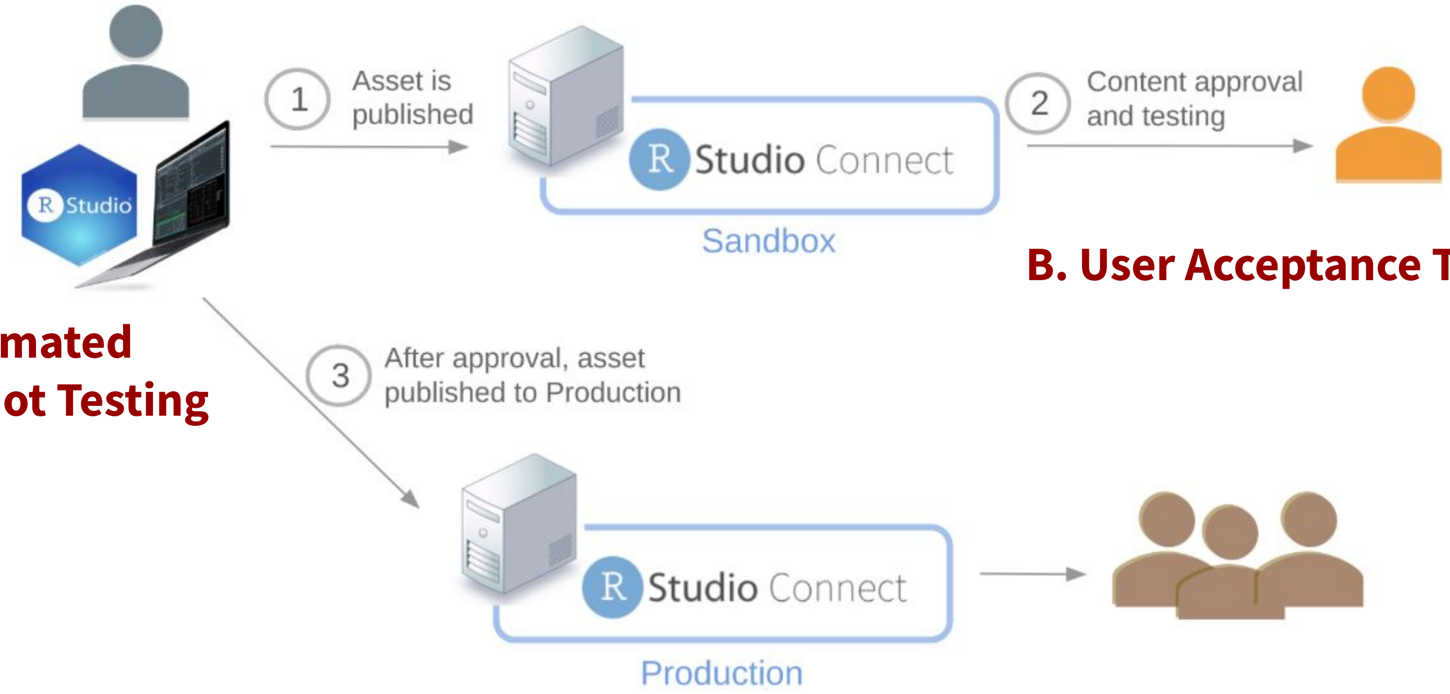
Reduce the risk of deploying a breaking change

Shared Goal: Automated Testing!

The improvement of daily work

Shared Goal: Getting a Sandbox!

Shorten the distance between development and production



A. Automated Snapshot Testing

B. User Acceptance Testing

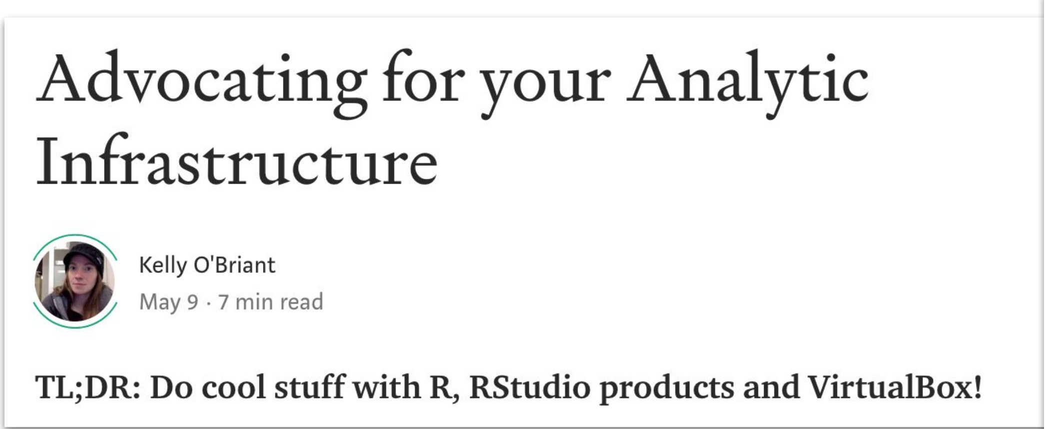
ADVOCATE FOR A SANDBOX PUBLISHING ENVIRONMENT

Shared Goal:

Shorten the distance between development and production

When developers begin to think of infrastructure as part of their application, stability and performance become normative.

- Jeff Geerling “Ansible for DevOps”

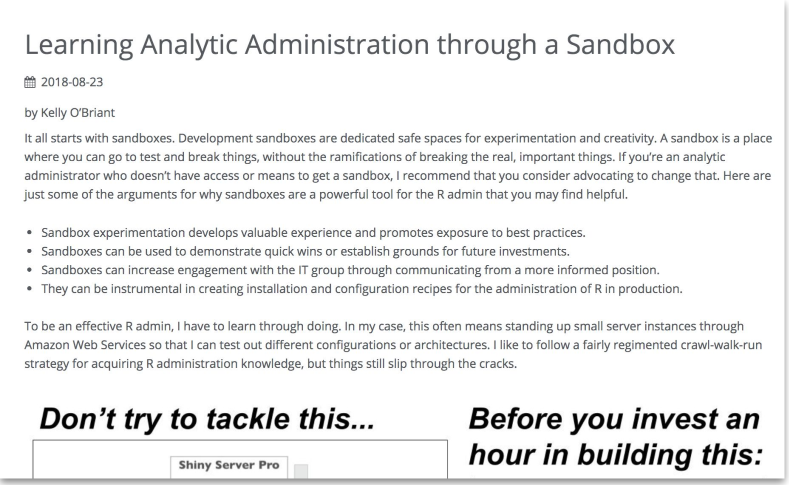


Advocating for your Analytic Infrastructure

Kelly O'Briant

May 9 · 7 min read

TL;DR: Do cool stuff with R, RStudio products and VirtualBox!



Learning Analytic Administration through a Sandbox

2018-08-23

by Kelly O'Briant

It all starts with sandboxes. Development sandboxes are dedicated safe spaces for experimentation and creativity. A sandbox is a place where you can go to test and break things, without the ramifications of breaking the real, important things. If you're an analytic administrator who doesn't have access or means to get a sandbox, I recommend that you consider advocating to change that. Here are just some of the arguments for why sandboxes are a powerful tool for the R admin that you may find helpful.

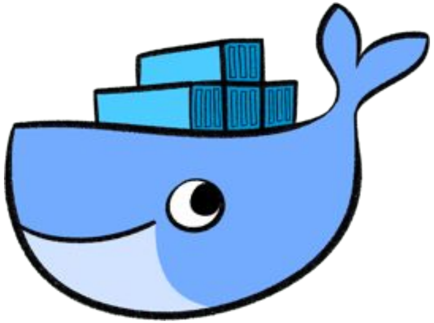
- Sandbox experimentation develops valuable experience and promotes exposure to best practices.
- Sandboxes can be used to demonstrate quick wins or establish grounds for future investments.
- Sandboxes can increase engagement with the IT group through communicating from a more informed position.
- They can be instrumental in creating installation and configuration recipes for the administration of R in production.

To be an effective R admin, I have to learn through doing. In my case, this often means standing up small server instances through Amazon Web Services so that I can test out different configurations or architectures. I like to follow a fairly regimented crawl-walk-run strategy for acquiring R administration knowledge, but things still slip through the cracks.

Don't try to tackle this...

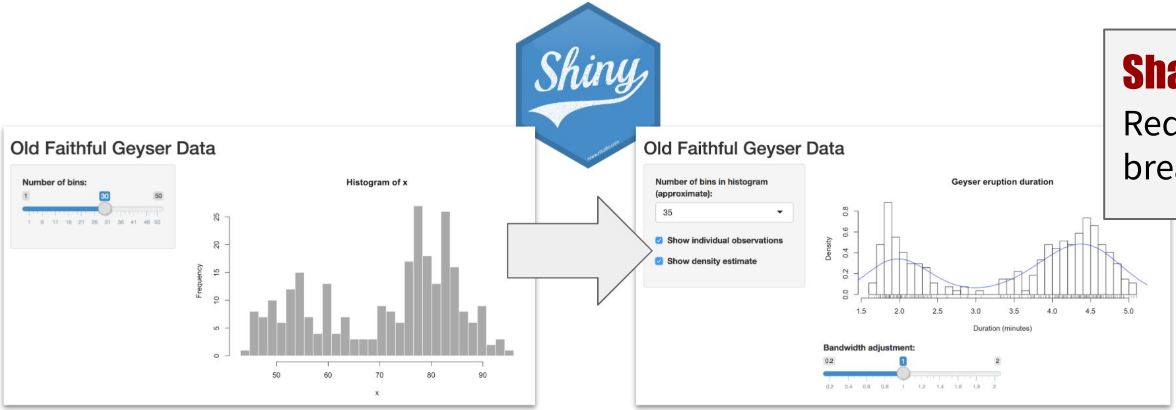
Before you invest an hour in building this:

Learning Environments for **Zero Dollars**



DevOps Learning: Decouple **deployment** from **release**

- **Deployment** is any push of code to an environment (test, prod)
- **Release** is when that code (feature) is made available to users



Old Faithful Geyser Data

Number of bins: 30

Old Faithful Geyser Data

Number of bins in histogram (approximate): 35

☑ Show individual observations

☑ Show density estimate

Bandwidth adjustment: 1

Version 1

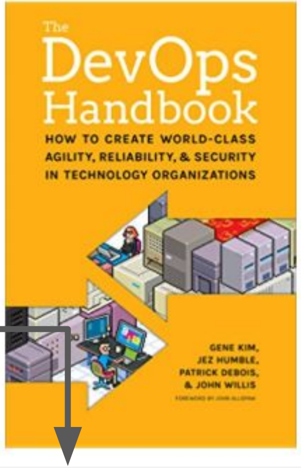
Version 2

Shared Goal:
Reduce the risk of deploying a breaking change!

Application-based release patterns vs. Environment-based release patterns

The DevOps Handbook

Three principles form the underpinnings of DevOps:


The DevOps Handbook
HOW TO CREATE WORLD-CLASS AGILITY, RELIABILITY, & SECURITY IN TECHNOLOGY ORGANIZATIONS
GENE KIM, JEZ HUMBLE, PATRICK DEBOIS, & JOHN WILLIS

1. Accelerate Flow

- Make work visible
- Limit Work in Progress (WIP)
- Reduce Batch Sizes
- Reduce the number of handoffs
- Continually identify and elevate constraints
- Eliminate hardships and waste

2. Utilize Feedback

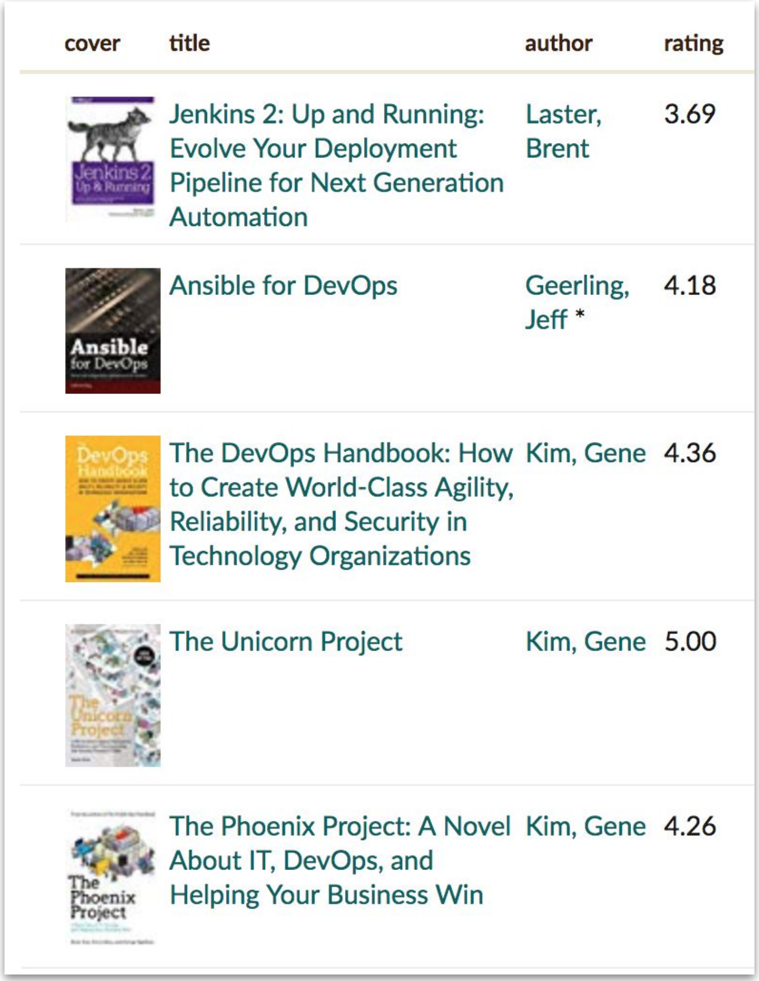
- See problems as they occur
- Swarm to solve problems and build new knowledge
- Keep pushing quality closer to the source
- Enable optimizing for downstream work centers

3. Learn and Experiment

- Enable organizational learning and a safety culture
- Institutionalize the improvement of daily work
- Transform local discoveries into global improvements
- Inject resilience patterns into daily work

More Books!

bit.ly/devops-bookshelf



cover	title	author	rating
	Jenkins 2: Up and Running: Evolve Your Deployment Pipeline for Next Generation Automation	Laster, Brent	3.69
	Ansible for DevOps	Geerling, Jeff *	4.18
	The DevOps Handbook: How to Create World-Class Agility, Reliability, and Security in Technology Organizations	Kim, Gene	4.36
	The Unicorn Project	Kim, Gene	5.00
	The Phoenix Project: A Novel About IT, DevOps, and Helping Your Business Win	Kim, Gene	4.26

Join the Community

How to be an empathetic witness:

- Listen
- Ask good questions
- Never judge
- Never advise* but offer your own experiences

(citation: Benjamin Hardy, PhD)

solutions.rstudio.com

community.rstudio.com

#radmins