PROD IS NOT A FOUR-LETTER WORD

How to confidently send R to production.

COLIN FAY
Colin FAY
Data Scientist & R-Hacker at ThinkR, a french company focused on Data Science & R Engineering.
Hyperactive open source developer.

http://rtask.thinkr.fr
http://twitter.com/_colinfay
http://github.com/colinfay
R & DATA SCIENCE ENGINEERING

TRAINING

CONSULTING
How to confidently send R to production.
What are we going to talk about today?
01. What does production even mean?
02. About R in production
03. Let’s talk about tools
04. R in production, the challenges
Have you ever heard?

"That’s good that we have this prototype in R, now let’s rewrite this with a production language"
We’ve come to a point that we’ve heard this so much that we convinced ourselves that this is true.
All programming languages are basically just bundles of trade-offs, and so no single language is going to be globally superior to another. There really are no production languages – only production engineers.

Gordon Shotwell

https://blog.shotwell.ca/posts/why_i_use_r/
DON’T TELL ME

R IS NOT PRODUCTION READY
DID YOU SAY ‘PRODUCTION’?
Question for machine learning people in industry: what does it mean for your work to be “in production” what qualifies your models to be “production ready”? See this a lot and I think the meanings change across different domains pretty quickly.

8:26 PM · 2 mars 2020 · Twitter for iPhone

https://twitter.com/jdetle
Production is anything that is run repeatedly (or continuously) and is **relied upon**.

*Mark Sellors*

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

*Joe Cheng*
As an engineer: Versioned, resilient, modular so it can be easily changed/iterated over and I can switch environments by configuration settings. More than business logic/prediction. Available/Integrated into a system. As anyone else: when it is producing value that makes impact.
IN PROD

IS USED
If nobody's using it, it's not in production

8:46 PM · 13 janv. 2020 · Twitter Web App

3 Retweets  13 J'aime
IN PROD

IS USED

IS RELIED UPON
If turning it off breaks your business, it’s in production

7:13 AM · 3 mars 2020 · Twitter for iPad

https://twitter.com/dierken
TO BE AVAILABLE

🍪 TO MAKE DECISION

🤖 TO DO THEIR DAILY JOB
TO KEEP RUNNING
TO BE BUG FREE
TO BE CUSTOMISABLE
TO BE SAFE TO UPDATE
IN PROD

- IS USED
- IS RELIED UPON
- HAS REAL LIFE IMPACT
En réponse à @peteskomoroch @jdetle et 9 autres

I’d explain the main difference as one of accountability. If someone can blame you when something goes wrong and you’re responsible for fixing in a timely manner, then it’s in production. The proxy I’d use is if you have an oncall rotation.

6:59 AM · 3 mars 2020 · Twitter for iPhone
WHAT CAN GO WRONG

❌ SOFTWARE IS UNAVAILABLE

🚨 UPDATE BREAKS THE CODE-BASE

🚫 UPDATE BREAKS ANOTHER SERVICE

جا RESULTS ARE NOT RELIABLE
REAL LIFE IMPACT

- DATA LOSS
- MONEY LOSS
- TIME LOSS
- BAD DECISIONS
> PRODUCTION_TOOLS == RELIABLE_TOOLS
[1] TRUE
ABOUT R AND PRODUCTION
How many ways are there to build a package in R?
RELY ON FRAMEWORKS
DON’T REINVENT THE WHEEL
TUTORIALS ARE EVERYWHERE
COMMENT, TEST, AND MONITOR
CONVENTIONS ARE AWESOME

(And so is the R community)
03 LET'S TALK ABOUT TOOLS
R PACKAGES

🤝 MAIN GOAL IS TO BE SHARED AND REUSED

🛠 YEARS OF TOOLKITS

(IT JUST A LIBRARY() AWAY)
DEVELOP

AUTOMATE

DOCUMENT

TEST
SHINY APPS WITH {GOLEM}

🛠 FRAMEWORK

🤖 AUTOMATION

🧠 COLLECTIVE INTELLIGENCE

https://www.golemverse.org/
golem 0.2.1

ColinFay released this yesterday

This release matches the CRAN release of \{golem\} v 0.2.1

Assets 2

- Source code (zip)
- Source code (tar.gz)

https://cran.r-project.org/package=golem
DEPLOY

📽 RSTUDIO

🐳 DOCKER / KUBE

🐱 AWS
R IN PRODUCTION
THE CHALLENGES
NOT THE LANGUAGE

NOT THE TOOLS
CULTURAL
ORGANISATIONAL

CHALLENGES
Believe it or not, I think the main problem is culture, not tech. People love building models but hate paying the interest payments on their tech debt.
Who has ever experienced?

"Woaw, your PoC is awesome, let’s put that on the production server and give access to 100 users!"
The good thing about R is that one can start using it and get results in a couple of hours.

The bad thing about R is that one can start using it and get results in a couple of hours.
R developers don’t learn R as software engineers.

R developers don’t know what’s needed for productions.

IT might not be receptive.
ADOPT A PROD MINDSET

🚨 NO MORE “THAT WORKS ON MY MACHINE”

📦 EVERYTHING IS A PACKAGE

👨‍⚖️ TEST AND DOCUMENT

✔️ DON’T START WITH TECHNICAL DEBT
👩‍💻 IT ARE YOUR FRIENDS
💪 INFRASTRUCTURE IS CRUCIAL
✍ RELIABLE CODE MATTERS
YOU CAN BE A GOOD SOFTWARE ENGINEER
R WORKS IN PRODUCTION
THANK YOU!

Questions?

COLIN@THINKR.FR
HTTPS://TWITTER.COM/_COLINFAY
HTTPS://TWITTER.COM/THINKR_FR
HTTPS://RTASK.THINKR.FR
HTTPS://COLINFAY.ME