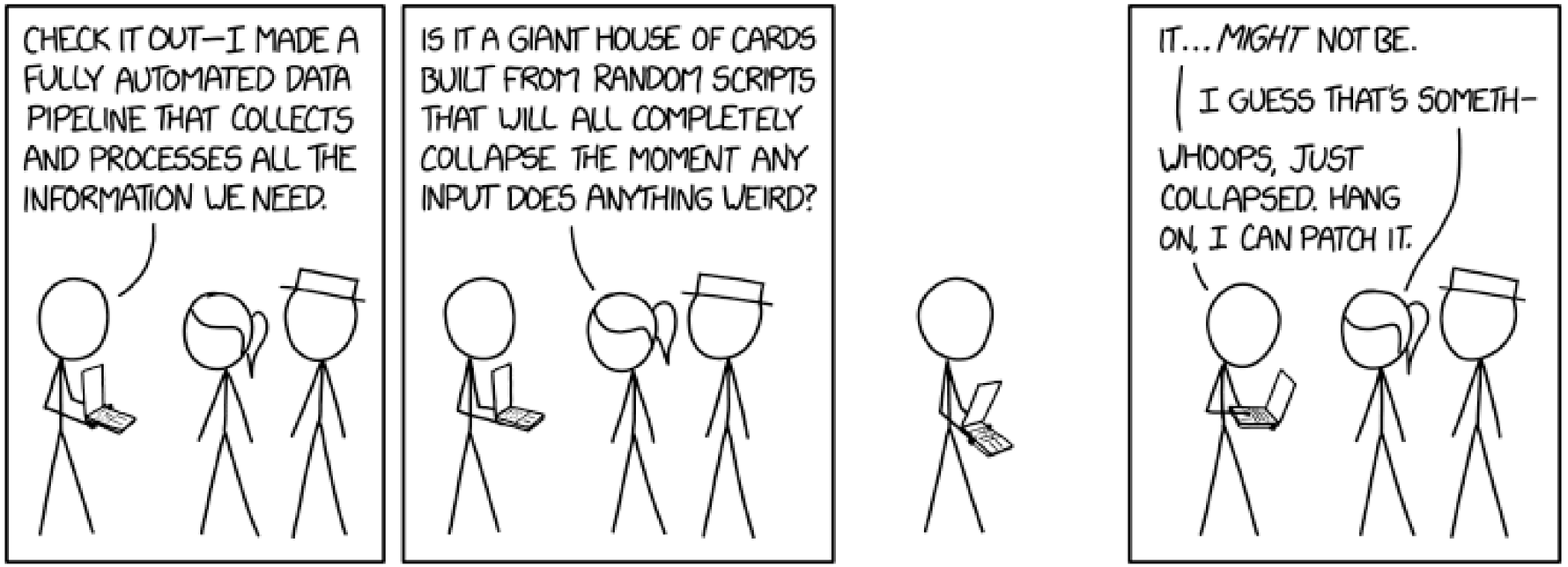
HOUSTON ANALYTICS

DEVOPS FOR DATA SCIENCE

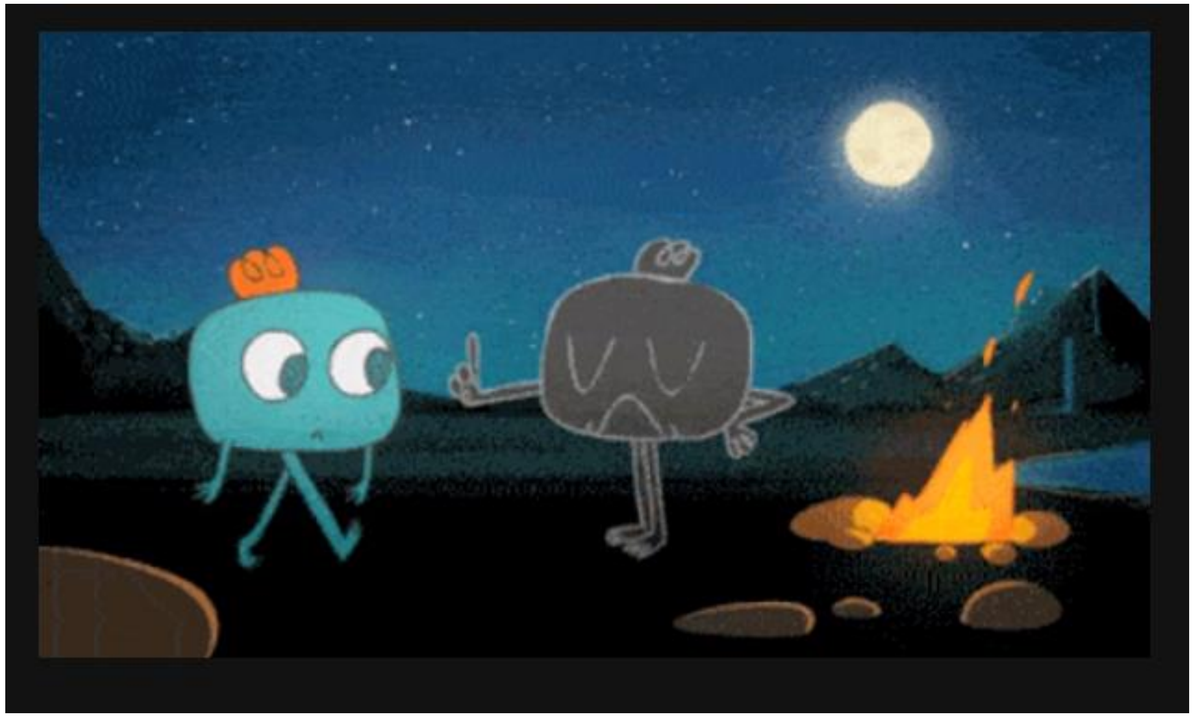
HOUSTON ANALYTICS



"IS THE PIPELINE LITERALLY RUNNING FROM YOUR LAPTOP?" "DON'T BE SILLY, MY LAPTOP DISCONNECTS FAR TOO OFTEN TO HOST A SERVICE WE RELY ON. IT'S RUNNING ON MY PHONE."

© HOUSTON ANALYTICS LTD

ABOUT ME



- HTTP://DEVOPSREACTIONS.TUMBLR.COM

- MSC IN STATISTICS 2003, PHD IN STATISTICS 2007, MINORS MATH AND CS
- 2008-2014 SENIOR RESEARCHER AT FINNISH FOREST RESEARCH INSTITUTE (METLA, NOW LUKE)
- 2015-2018 SENIOR APPLICATION SPECIALIST AT CSC – IT CENTER FOR SCIENCE
- AUGUST 2018 – PRESENT SENIOR DATA SCIENTIST AT HOUSTON ANALYTICS
- R USER SINCE 1999

© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

DevOps for Data Scientists: Taming the Unicorn – Towards Data Science

https://towardsdatascience.com/devops-for-data-scientists-taming-the... ▾ Käännä tämä sivu

1.7.2018 - When most **data scientists** start working, they are equipped with all the neat math concepts they learned from school textbooks. However, pretty ...

The Growing Significance Of DevOps For Data Science - Forbes

https://www.forbes.com/.../the-growing-significance-of-devops-for-... ▾ Käännä tämä sivu

4.11.2018 - Machine learning brings a new dimension to **DevOps**. Along with developers, operators will have to collaborate with **data scientists** and data ...

Q&A: How to Use DevOps for Data Science - DataScience.com

https://www.datascience.com/blog/devops-data-science ▾ Käännä tämä sivu

29.9.2017 - To understand why larger companies are embracing a **DevOps** mentality, we sat down with Pam McCaslin, **data scientist** and principal **DevOps** ...

DevOps for Data Scientists - Lynda.com

https://www.lynda.com/Data-Science.../DevOps-Data-Scientists/5780... ▾ Käännä tämä sivu

10.5.2018 - Learn the principles of supporting **DevOps** and how to apply them to **data science**. Kävit tällä sivulla 11.1.2019.

© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

WHAT IS DEVOPS, ANYWAY?

WIKI DEFINITION:
" **DEVOPS** (A CLIPPED COMPOUND OF "DEVELOPMENT" AND "OPERATIONS") IS A SOFTWARE DEVELOPMENT METHODOLOGY THAT COMBINES SOFTWARE DEVELOPMENT (*DEV*) WITH INFORMATION TECHNOLOGY OPERATIONS (*OPS*). THE GOAL OF DEVOPS IS TO SHORTEN THE SYSTEMS DEVELOPMENT LIFE CYCLE WHILE DELIVERING FEATURES, FIXES, AND UPDATES FREQUENTLY IN CLOSE ALIGNMENT WITH BUSINESS OBJECTIVES"

MY FRIEND PATRIC:
"DEVOPS TO ME IS THE POLITE TERM FOR A CORPORATE CLUSTERF**K IN WHICH THE 'DEVOPS GUYS' DO EVERYTHING WHILE EVERYBODY ELSE IS IN MEETINGS DISCUSSING THE SAID GUYS PRODUCTIVITY. [XXX] IS DOING IT AND A COUPLE OF YEARS AGO I HAD SOME INTERVIEWS WITH THEM, BUT WHEN I HEARD THEIR IDEA OF DEVOPS, I DECLINED. SOMEHOW, PROGRAMMING AND ATTENDING 24/7 TO SERVERS ISN'T EXACTLY MY IDEA OF A BALANCED LIFE."

© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

”IN PRODUCTION”

BY ED AND EDDIE FROM PALO ALTO, USA - SFOPERAHOUSE2.JPGUPLOADED BY MAYBEMAYBEMAYBE, CC BY-SA 2.0, HTTPS://COMMONS.WIKIMEDIA.ORG/W/INDEX.PHP?CURID=22830509

- A **SCRIPT** IS NOT THE WHOLE SHOW!
- NOVELIST IS NOT THE SAME THING AS PLAYWRIGHT
- FOR A FULL THEATRICAL SHOW YOU NEED A STAGE, SOME ACTORS, AND MOST IMPORTANTLY AN AUDIENCE
- IN THEATRE THERE USUALLY IS A SPECIAL ROLE CALLED A PRODUCER WHO PUTS ALL THE PIECES TOGETHER

NOTE: IN PRODUCTION YOU SHOULD PROBABLY NOT HAVE LONG SCRIPTS OF FREE FLOATING CODE. USE FUNCTIONS. CONSIDER CREATING AN ACTUAL R PACKAGE.

© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

VERSION CONTROL (GIT)

- ALSO KNOWN AS SOURCE CONTROL
- GIT IS NOT THE ONLY VERSION CONTROLS SYSTEM (BUT LET'S FACE IT...)

- A BACKUP SYSTEM AND A TIME MACHINE
- MAKES COLLABORATION POSSIBLE
- IN PRACTICE: IT IS A THING THAT RESIDES IN A FOLDER IN YOUR FILE SYSTEM, TURNING IT IN TO A *REPOSITORY*. YOU TAKE SNAPSHOTS OF THE STATE OF THE WHOLE REPOSITORY, AND GIT KEEPS TRACK OF THEM, EVEN ACROSS MANY SIMULTANEOUS AND ALTERNATIVE CHANGES (VIA BRANCHES)
- FOR STARTING TO LEARN CONSIDER THE LINK IN RSTUDIO, OR HTTPS://CODEREFINERY.ORG/

HTTPS://XKCD.COM/1597/

"IF THAT DOESN'T FIX IT, GIT.TXT CONTAINS THE PHONE NUMBER OF A FRIEND OF MINE WHO UNDERSTANDS GIT. JUST WAIT THROUGH A FEW MINUTES OF 'IT'S REALLY PRETTY SIMPLE, JUST THINK OF BRANCHES AS...' AND EVENTUALLY YOU'LL LEARN THE COMMANDS THAT WILL FIX EVERYTHING."

24.1.2019

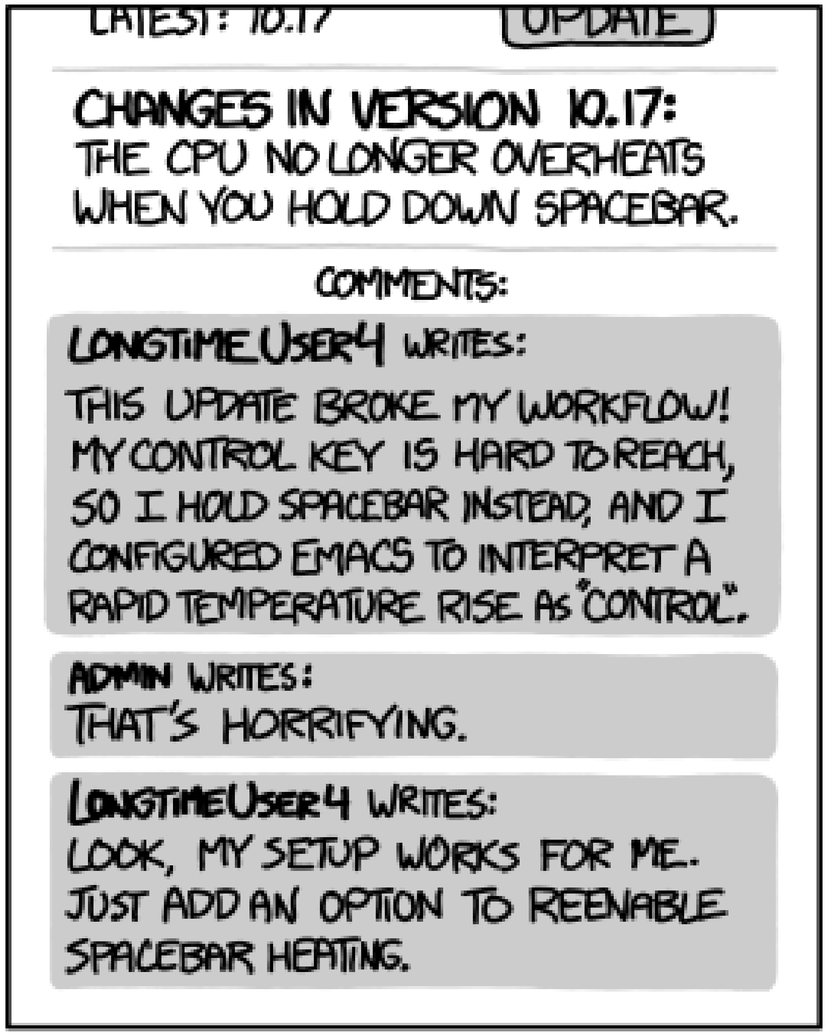
© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

(AUTOMATIC) TESTING

- WRITING AND RUNNING *UNIT TESTS* WILL HELP YOU NOT BREAK YOUR CODE
- **"BUT WHY WOULD I BREAK MY CODE??"**
- FOR EXAMPLE TRAVIS CI ON GITHUB WILL RUN AUTOMATIC TESTS
- JENKINS IS ANOTHER TOOL (AND BOTH DO MORE THAN JUST TESTS)
- FOR R THERE IS TESTTHAT PACKAGE AND RSTUDIO

- (DO NOT BE CONFUSED BY REGRESSION TESTS!)



EVERY CHANGE BREAKS SOMEONE'S WORKFLOW.

© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

CLOUD COMPUTING

- CLOUD COMPUTING IS NOWADAYS EASY AND CHEAP. SO NOW EVERYONE IS THEIR OWN SYSADMIN (AND DEV HAS BECOME DEVOPS)
- TOOLS FOR AUTOMATING: ANSIBLE, PUPPET, TERRAFORM (INFRASTRUCTURE AS CODE, IAC)
- VIRTUAL MACHINES?
- CONTAINERS? / DOCKER?? / KUBERNETES???
- ONE PRACTICAL POINT OF VIEW: CONTAINERS MAKE IT EASY FOR YOU TO RUN SEVERAL VERSIONS OF YOUR APPLICATION AT THE SAME TIME

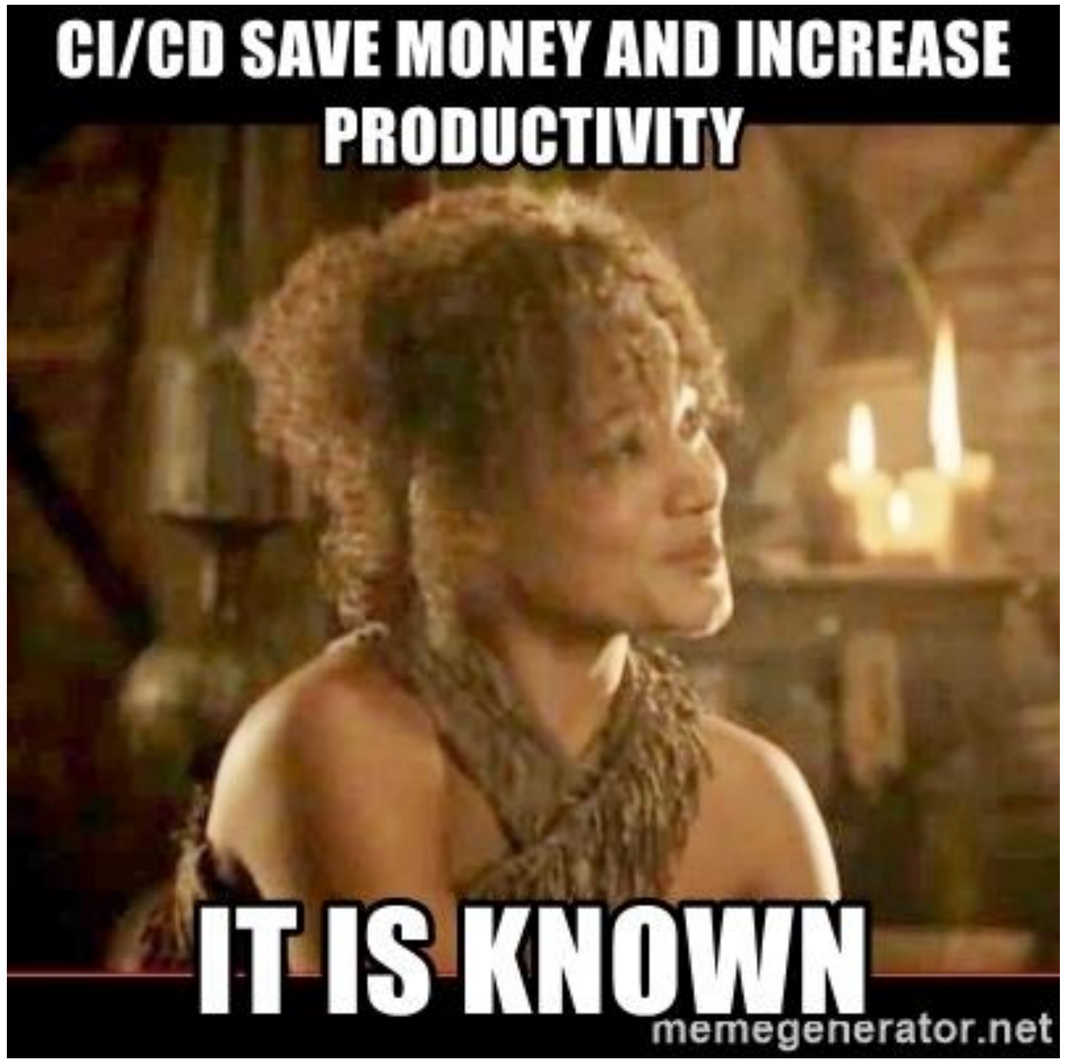
"EVERYBODY HAS A TESTING ENVIRONMENT. SOME PEOPLE ARE LUCKY ENOUGH THAT THEY HAVE A SEPARATE ENVIRONMENT FOR RUNNING PRODUCTION."

© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

CI / CD

- CONTINUOUS INTEGRATION, CONTINUOUS DEPLOYMENT (OR WAS IT DELIVERY…?)
- THE POINT IS: YOU WANT THE END USERS OF YOUR APPLICATION TO GET (STABLE!) IMPROVEMENTS ALL THE TIME

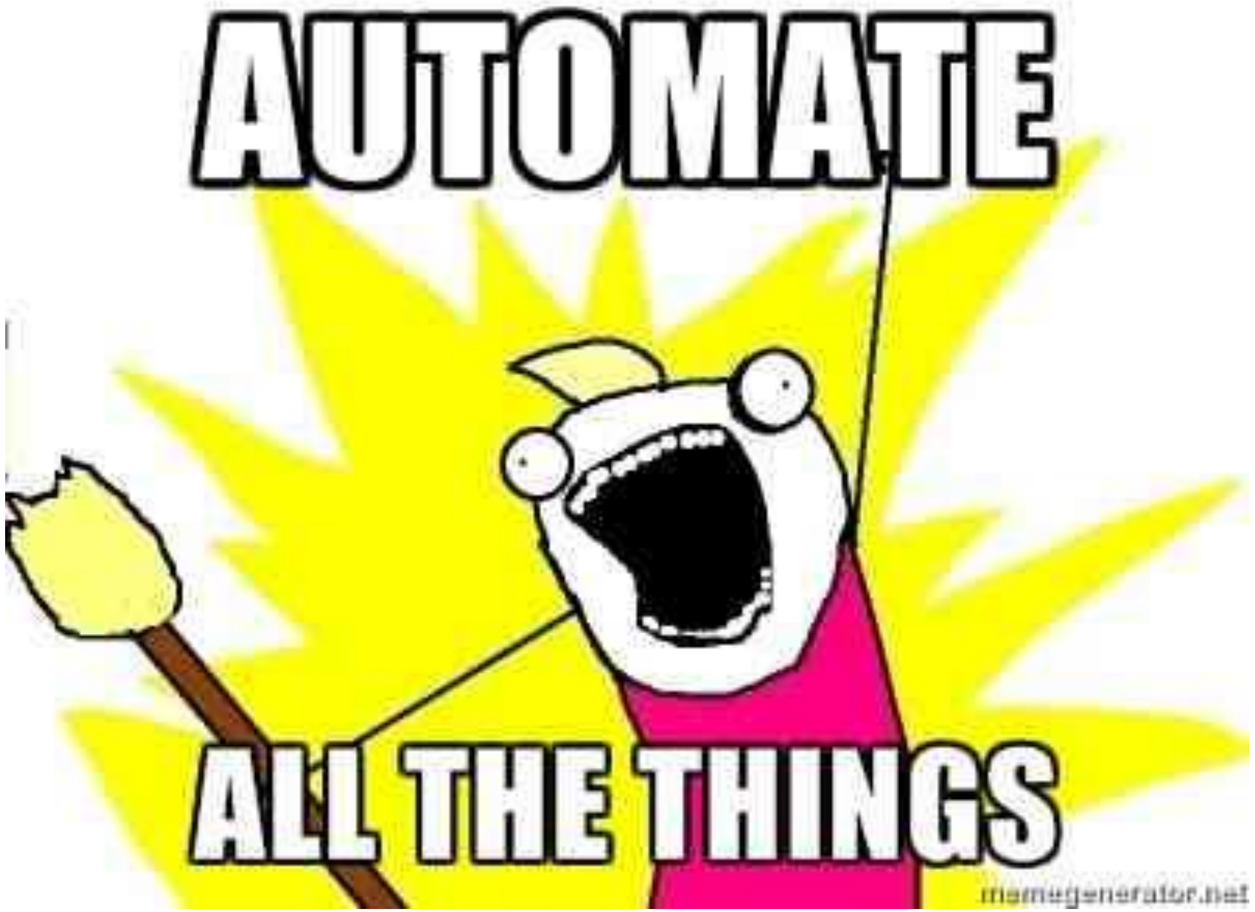


© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

A BRIEF RECAP

- AS A DATA SCIENTIST:
- BE AWARE THAT YOU ACTUALLY ARE A DEVELOPER
- THAT YOU ARE HANDLING YOUR DEVELOPMENT LIFE CYCLE *SOMEHOW* ANYWAY SO BETTER GIVE IT SOME THOUGHT
- THAT YOUR APPLICATION WILL BE RUNNING SOMEWHERE SO BETTER GIVE THAT SOME THOUGHT TOO
- MAKE YOUR CODE PERISHABLE. MAKE YOUR ENVIRONMENT PERISHABLE. AUTOMATE (MORE) THINGS.

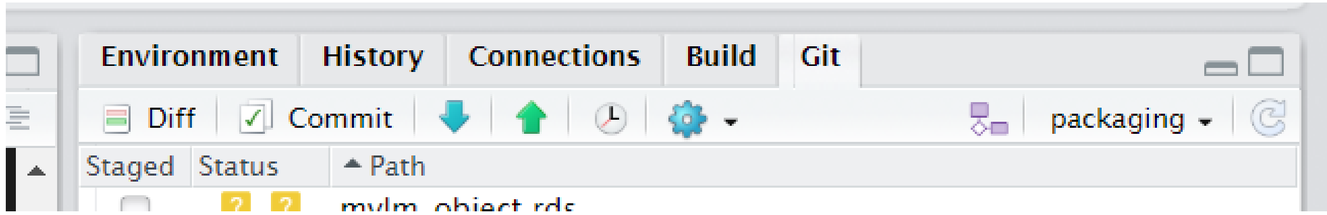


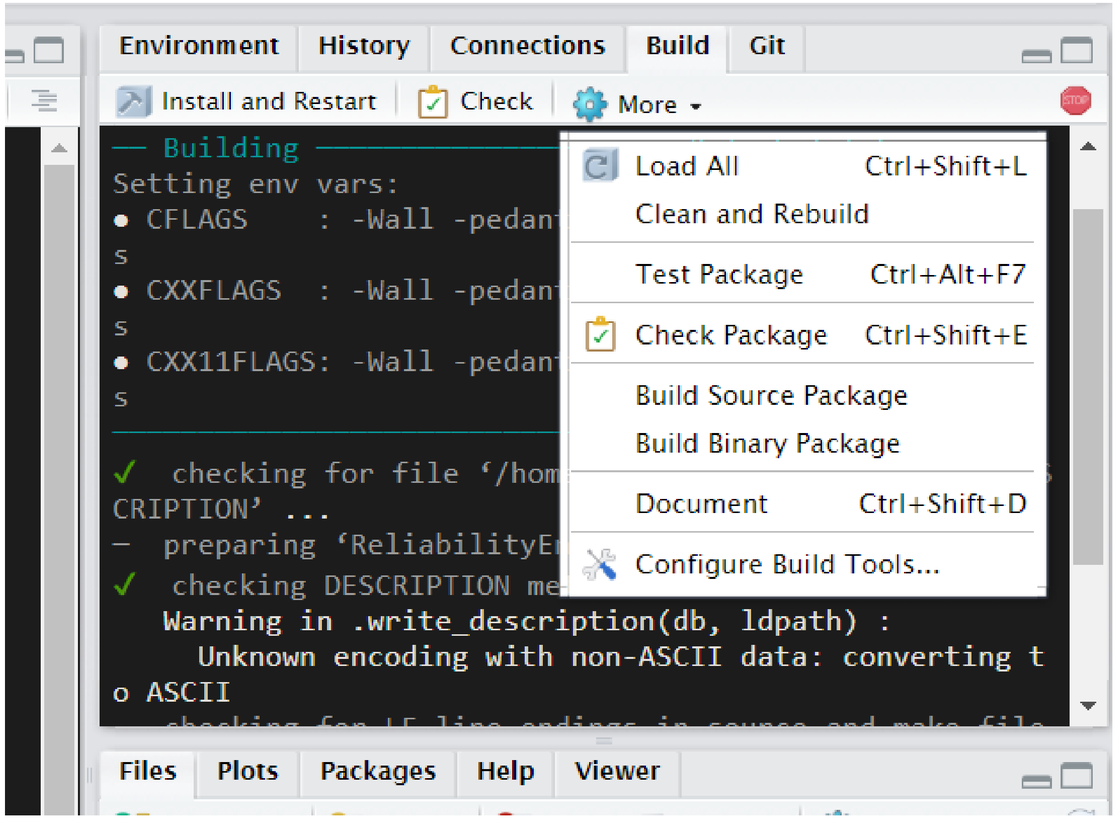
© HOUSTON ANALYTICS LTD

HOUSTON ANALYTICS

WHAT YOU CAN DO WITH RSTUDIO

- USE PROJECTS. *SERIOUSLY. I MEAN IT.*
- RSTUDIO HAS A VERY NICE GIT INTERFACE. EASY WAY TO START.
- GIVE YOUR CODEBASE SOME STRUCTURE (NO MORE SPAGHETTI CODE)
- IN FACT, CONSIDER MAKING A PACKAGE OUT OF YOUR CODEBASE
- IF YOU PLAN TO CONTAINERIZE, PACKAGING BECOMES ALMOST A NECESSITY (OR CONSIDER SWITCHING TO PYTHON…)





SPEAKING OF PYTHON… ANYONE INTERESTED IN A PYTHON DATA SCIENCE MEETUP? WILLING TO SPEAK? KNOW SOMEONE WHO IS?

© HOUSTON ANALYTICS LTD

SEIJA SIRKIÄ

SEIJA.SIRKIA@HOUSTON-ANALYTICS.COM
+358 40 5858509