Analytical Pipelines - Tools and Culture

Andrea Schnell
eRum 2018
Environment

ls()

Office


IT

## [1] closed up IT infrastructure very rigid firewall

Users

## [1] Excel natives R training «on-the-job»
Workflow

Today

- Data store
- Statistical Analysis
- Visualizations
- Text
- Publication

Tomorrow: Pipeline
Mission
Reproducible Analytical Pipelines

GOV.UK Data in Government

Transforming the process of producing official statistics

Reproducible Analytical Pipelines

RAP Companion
How to get there?

Tools

Culture
Tools

Data storage

Data base

```r
library(ROracle)
dbReadTable()
dbWriteTable()
```

- traceable
- secure
- less redundancy

- how to
- error handling (encoding, time zone)
Tools

Corporate Design Toolbox

```r
library(statR)
```

Colors, ggplot theme

RMardown template
Tools
Next steps

R

- RProjects
- RPackagess

Dependencies

- docker is too heavy
- packrat vs. checkpoint
- is drake an option?

Infrastructure

- RStudio Server
- Git/Github
Tools
Resistance

No thanks!

We are too busy
It's the Culture, Stupid!
Culture
Show & Tell

In person: Internal Meetups

- Solve problems
- New R Packages
- Ideas & inspiration

Virtual: Forum

- How To
- Code snippets
- Q & A
Culture
Community
«The use of R in public administration»

Talk 1: Read and process unstructured data with R
Max Grütter, Statistical Office
Slides

Talk 2: Mastering Corporate Design with R
Andrea Schnell and Thomas Lo Russo, Statistical Office
Slides

Talk 3: Illustration of pupils’ diversity with R
Flavian Imlig, Zurich State Department of Education
Slides

Talk 4: Making Public Data Public
Christoph Sax, Christoph Sax Data Analytics
Slides

https://statistikzh.github.io/RMeetup
Culture
Skills

+ • functions
  • Code review
  • Code style guide

- • copy & paste
  • loooooong R scripts
  • tool frictions
Reproducibility

Cooperation

Pace setting
Your experiences?

Thanks!

Andrea Schnell
andrea.schnell@statistik.ji.zh.ch

@schnllr

https://schnllr.github.io/eRum

Slides created via the R package xaringan.