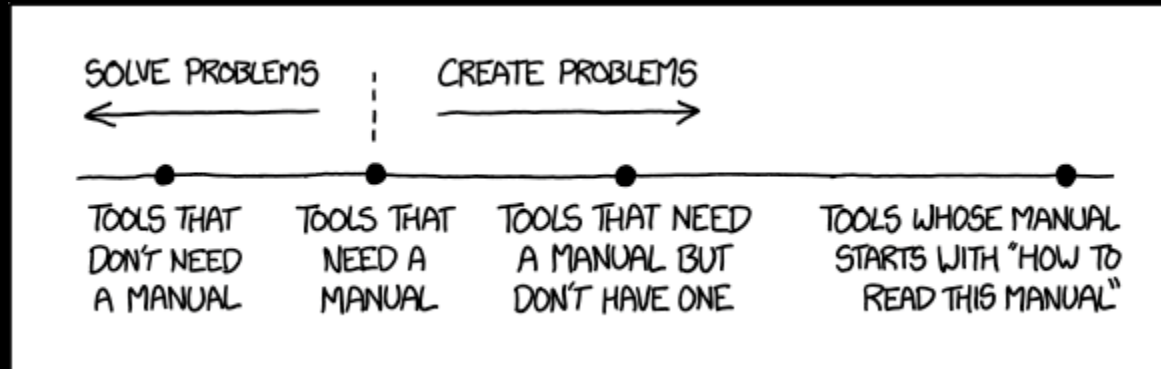


# Use The Right Tool



## R in Infrastructure and Business Analytics

Ágnes Salánki



# Use the right tool for the right job?



# Use the right tool for the right job?

ALWAYS BE YOURSELF!

ALWAYS USE THE RIGHT TOOL!



# Use the right tool for the right job?

ALWAYS BE YOURSELF!  
— UNLESS —  
YOU CAN BE  
BATMAN  
— THEN —  
ALWAYS BE BATMAN

ALWAYS USE THE RIGHT TOOL!



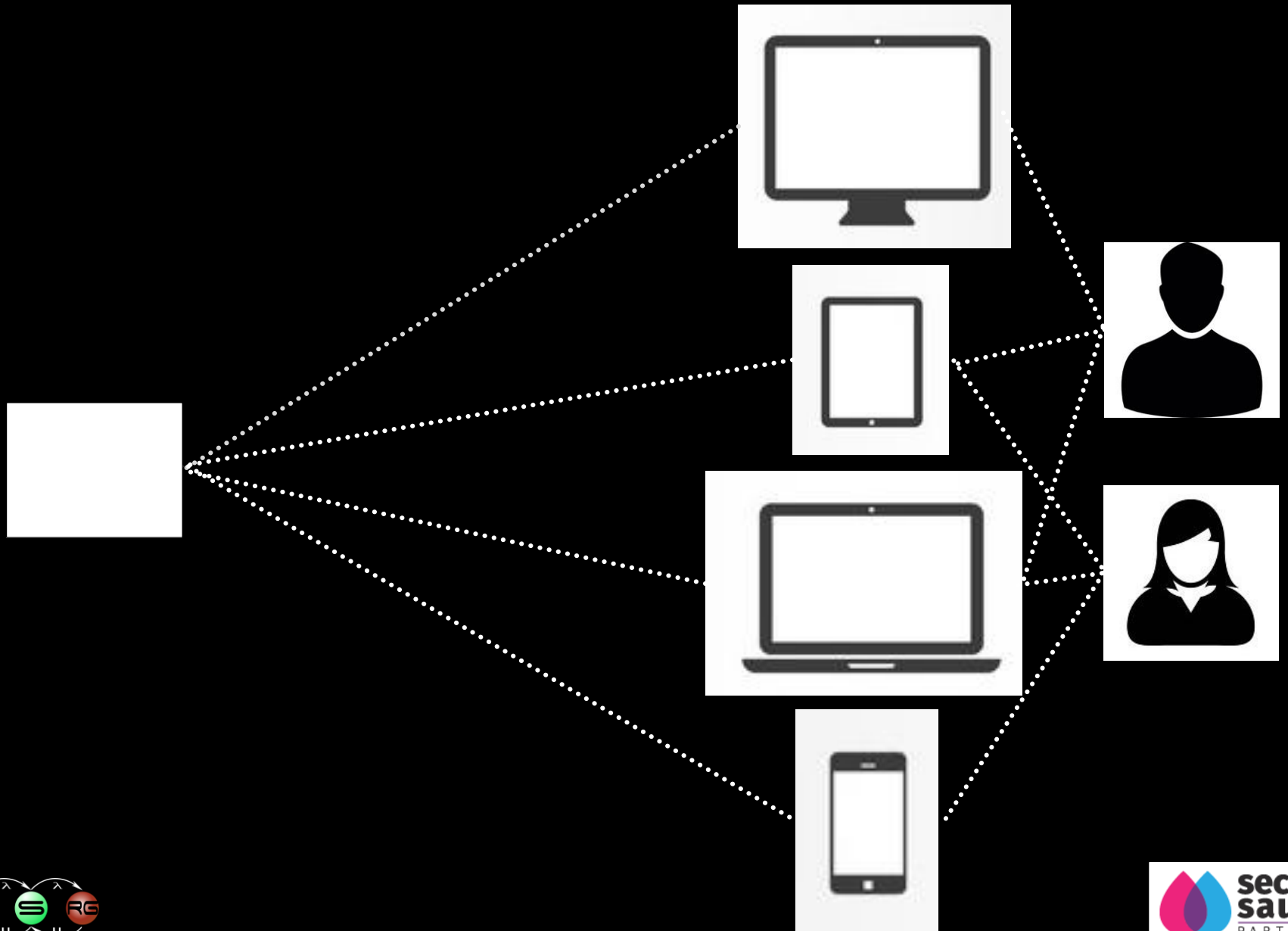
# Use the right tool for the right job?

ALWAYS BE YOURSELF!  
— UNLESS —  
YOU CAN BE  
BATMAN  
— THEN —  
ALWAYS BE BATMAN

ALWAYS USE THE RIGHT TOOL!  
— UNLESS —  
YOU KNOW ~~ONLY~~  
R  
— THEN —  
ALWAYS USE R

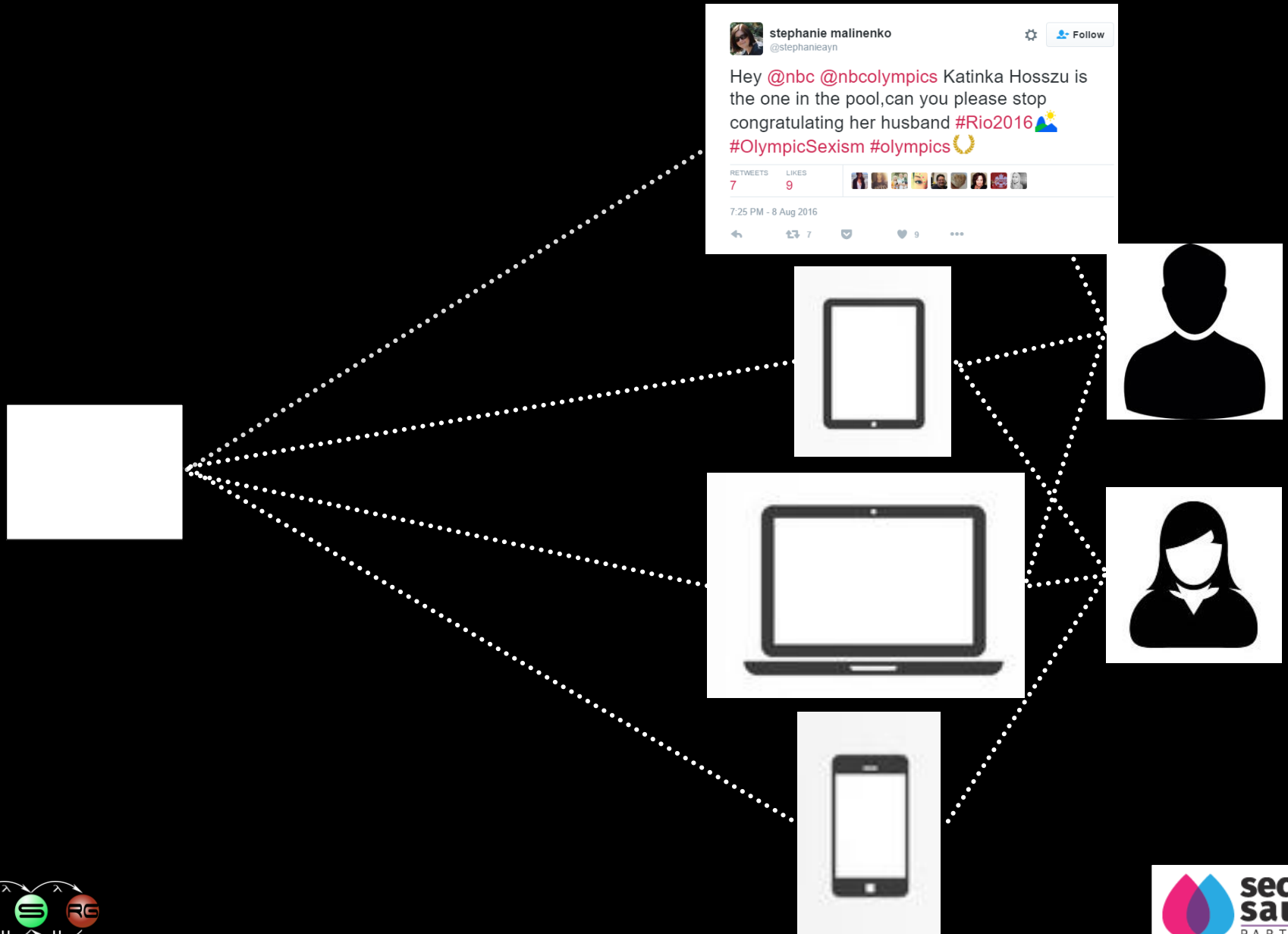


# Infrastructure vs. Business Analytics: Two sides of the same system



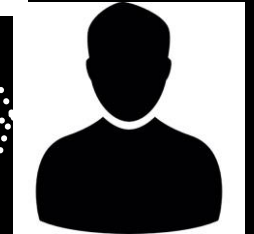
# Infrastructure vs. Business Analytics:

## Two sides of the same system



# Infrastructure vs. Business Analytics:

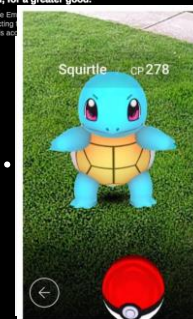
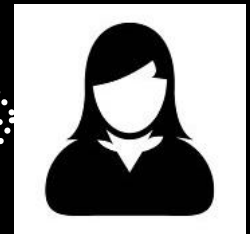
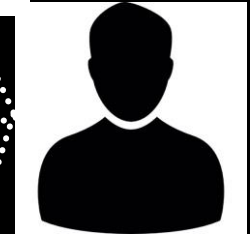
## Two sides of the same system





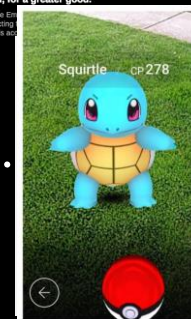
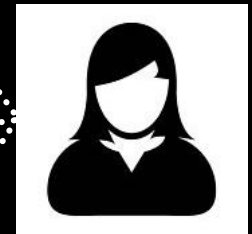
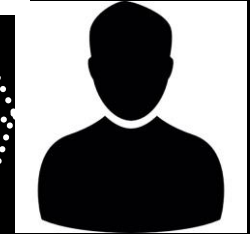
# Infrastructure vs. Business Analytics:

## Two sides of the same system



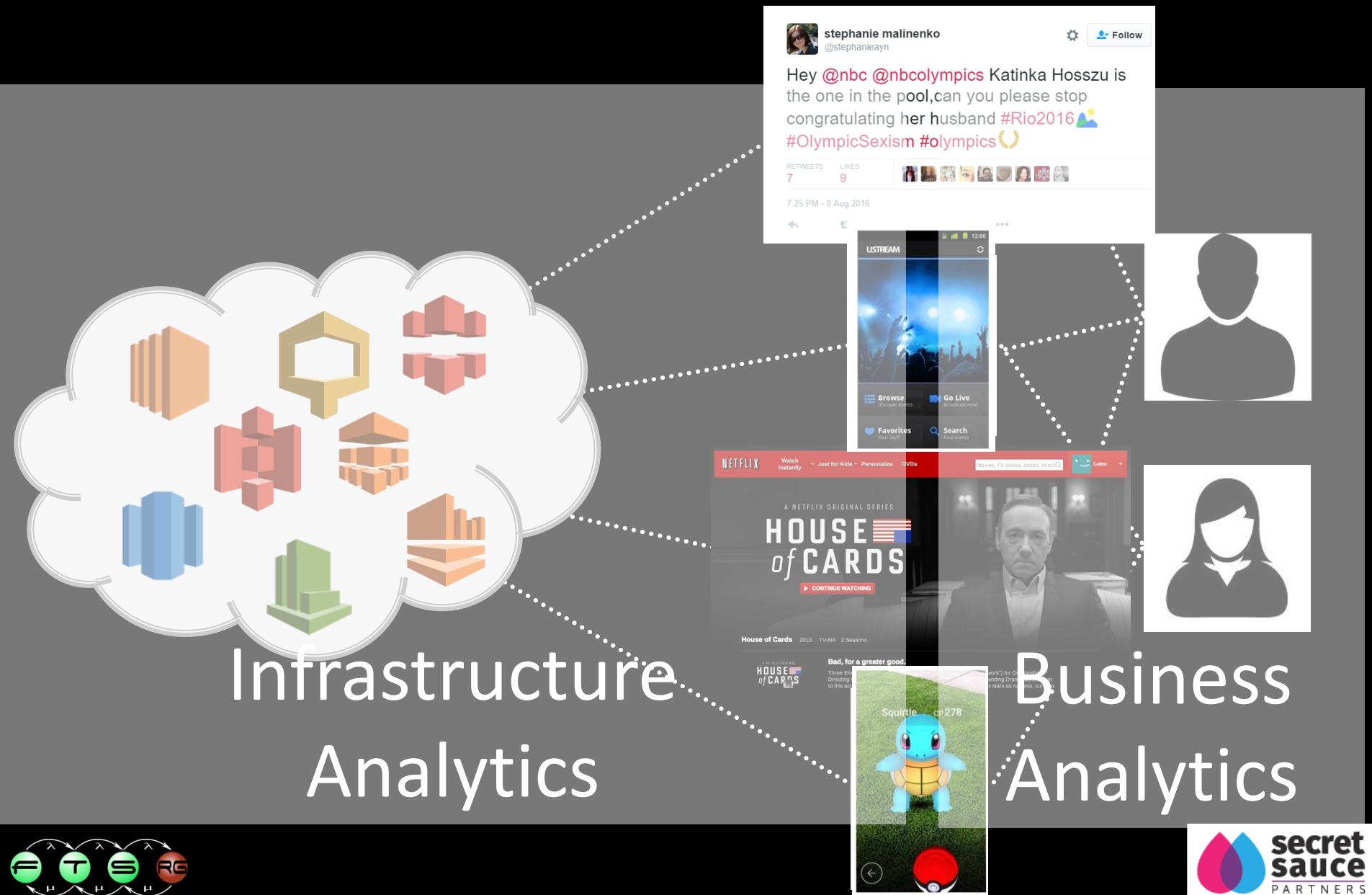
# Infrastructure vs. Business Analytics:

## Two sides of the same system

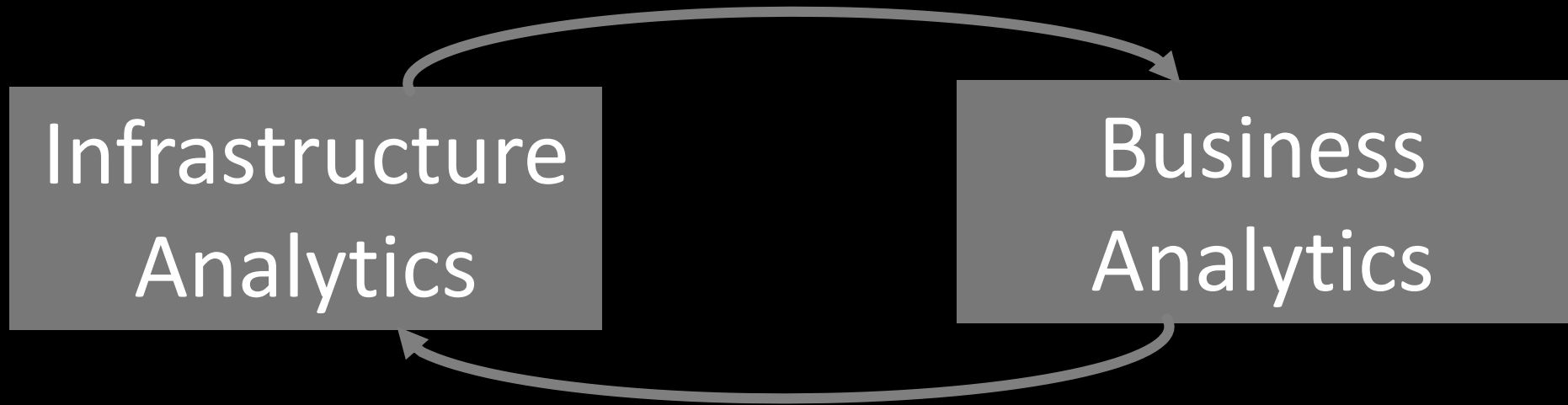


# Infrastructure vs. Business Analytics:

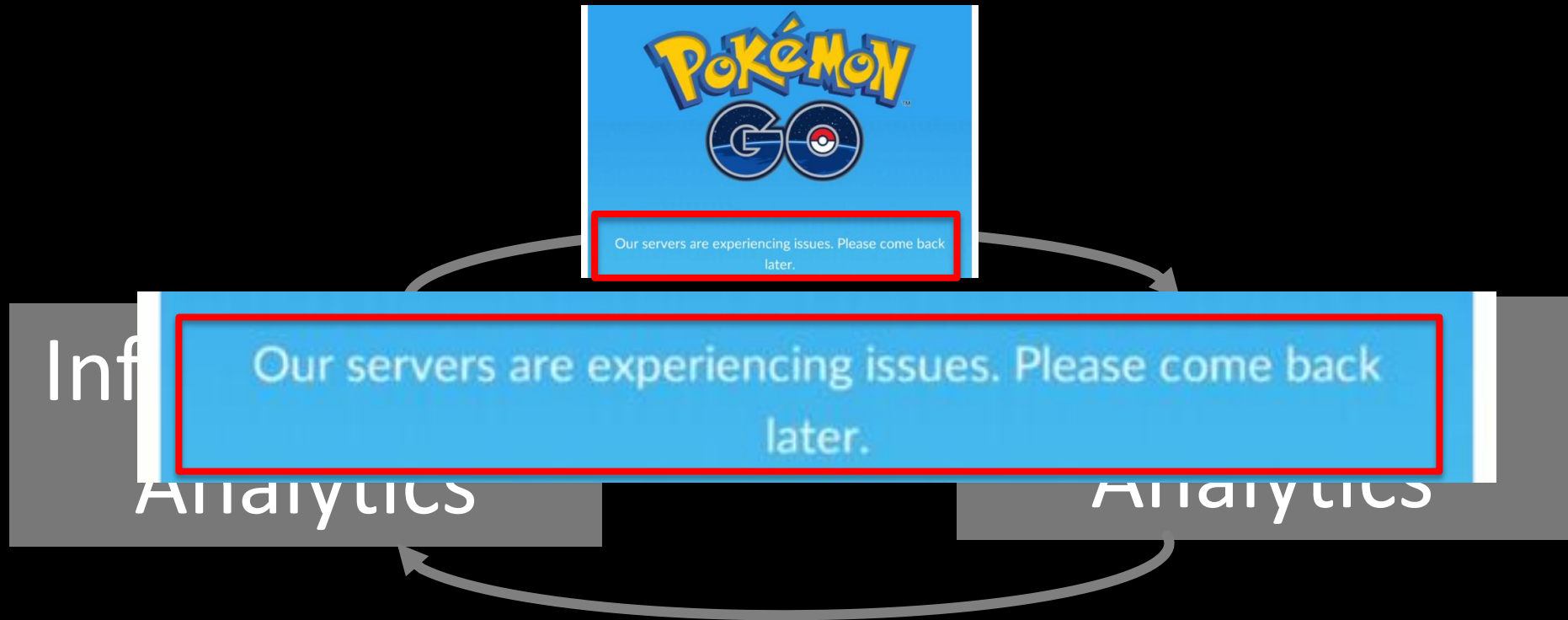
## Two sides of the same system



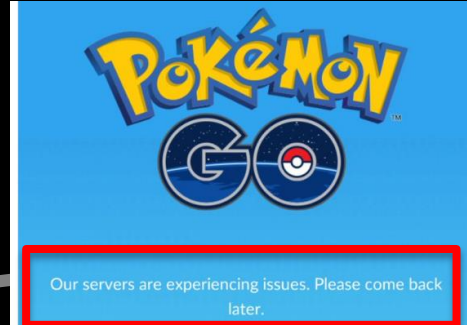
# Infrastructure vs. Business Analytics: Two sides of the same system



# Infrastructure vs. Business Analytics: Two sides of the same system

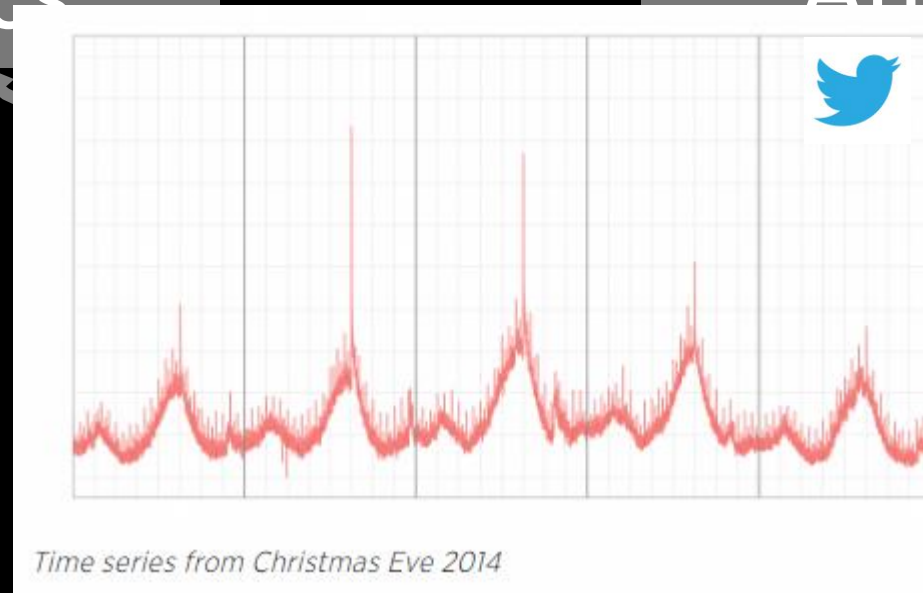


# Infrastructure vs. Business Analytics: Two sides of the same system

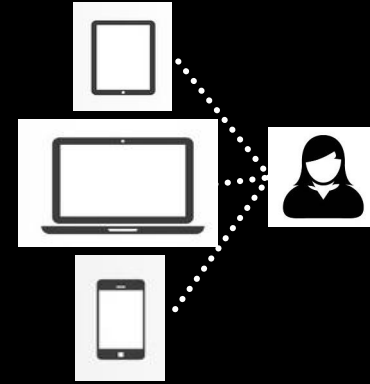


Infrastructure  
Analytics

Business  
Analytics



# My background



- PhD student
- Fault Tolerant Systems Research Group
  - Systems with availability of 99.99%
- Data Analyst in Product Management
- Secret Sauce Partners
  - Data-driven fashion industry

# My background



## JACQUARD LACE GOWN

CYNTHIA ROWLEY

USD\$ 698

### SIZE

0

2

4

**f** FIT PREDICTOR [calculate your size](#)

**COLOR** BLACK & BURGUNDY

**ADD TO BAG**

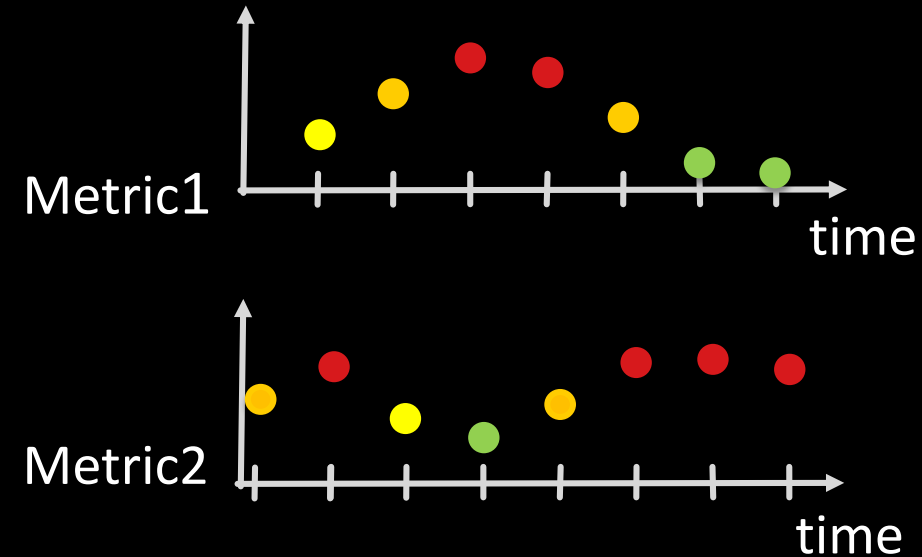
**VIEW FULL DETAILS**

**+** [Find Similar Items](#)

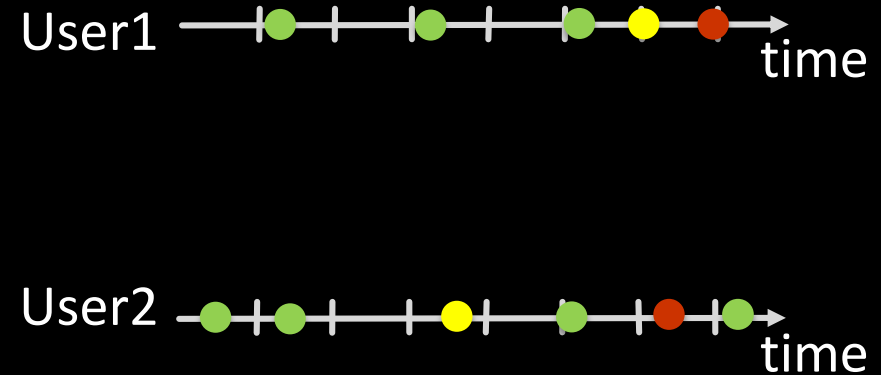


# Typical input

- Continuous sampling

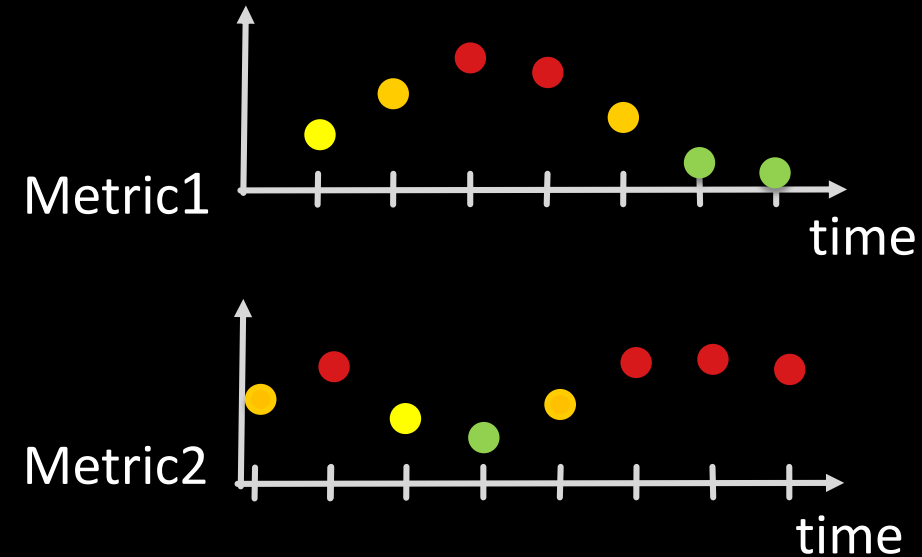


- Stream of events

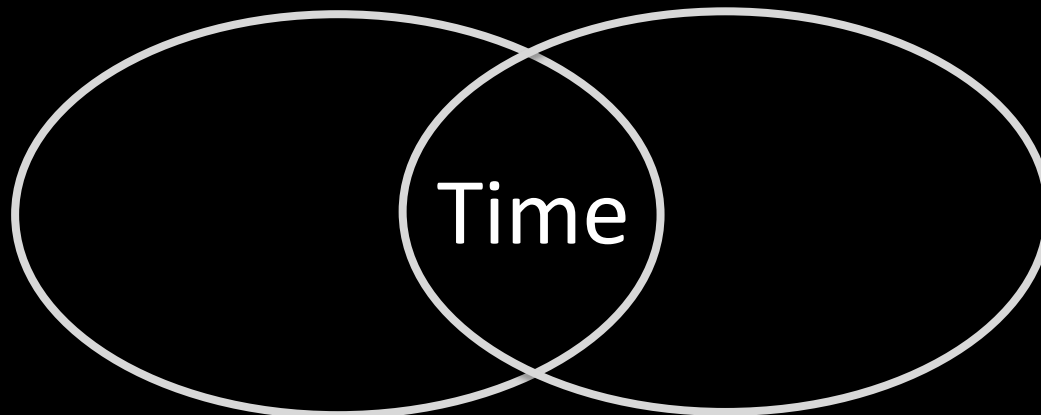
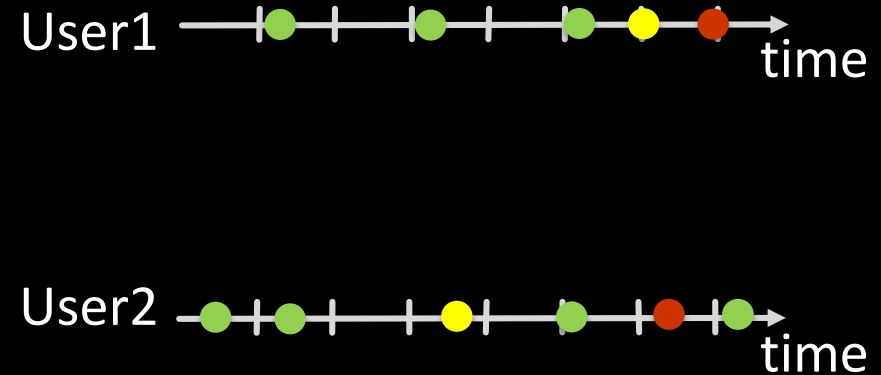


# Typical input

- Continuous sampling



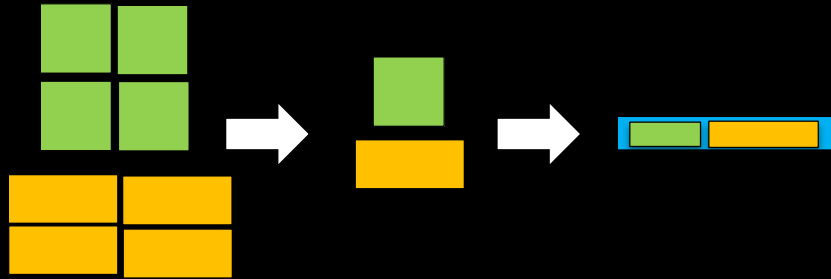
- Stream of events



# Getting the data

- Archived data: csv

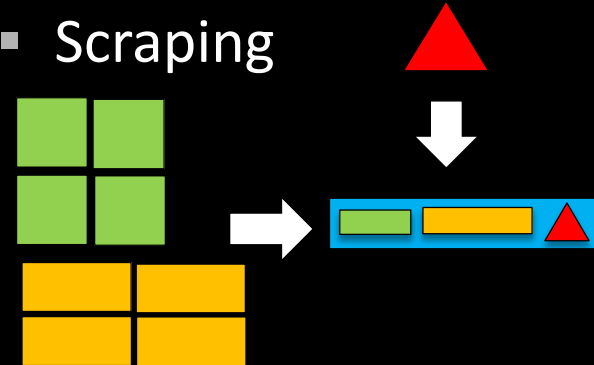
- Read-filter-join



- Archived data: databases

- Joins in SQL already

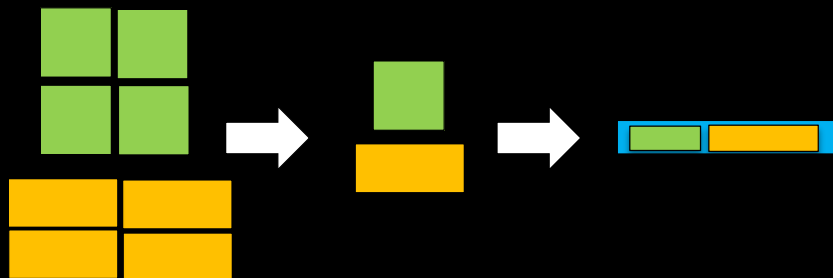
- Scraping



# Getting the data

- Archived data: csv

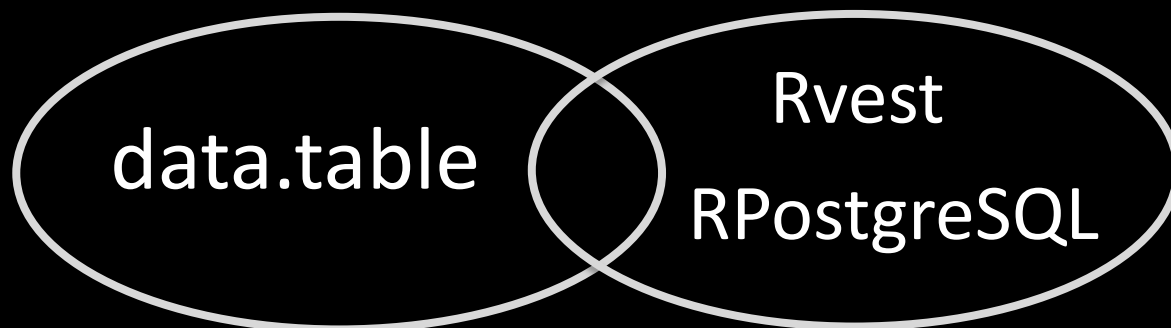
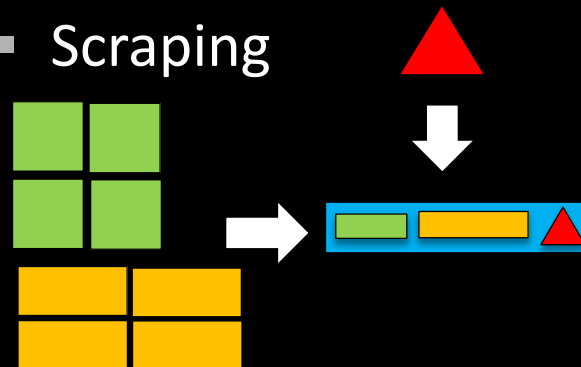
- Read-filter-join



- Archived data: databases

- Joins in SQL already

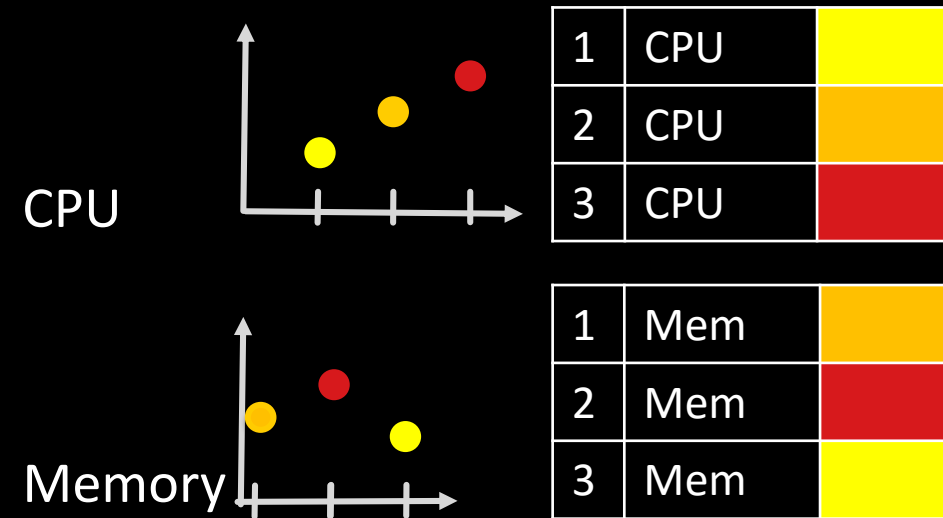
- Scraping



# Transforming the data

- Long format

- when, where, what, value



- Wide format

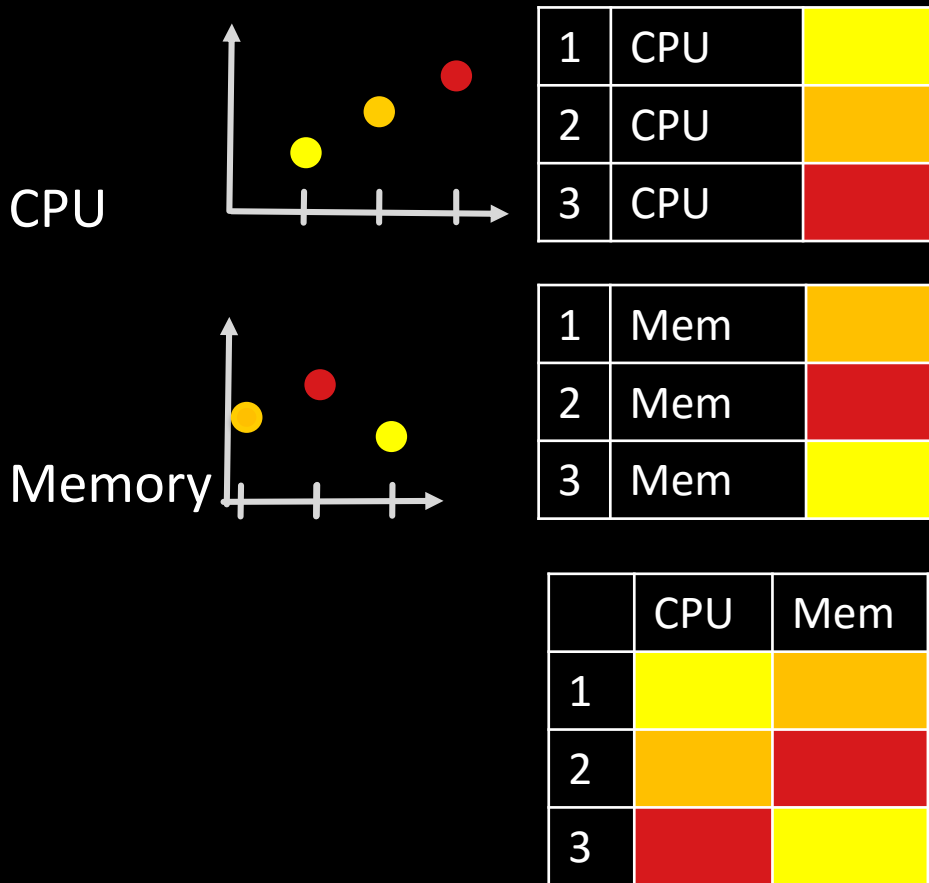
- lots of additional information



# Transforming the data

- Long format

- when, where, what, value



- Wide format

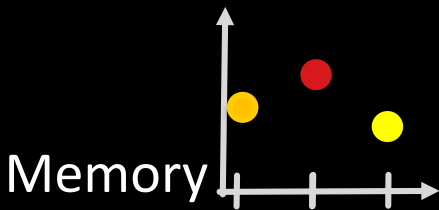
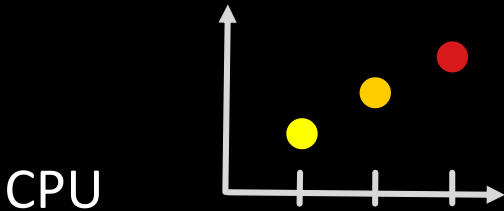
- lots of additional information



# Transforming the data

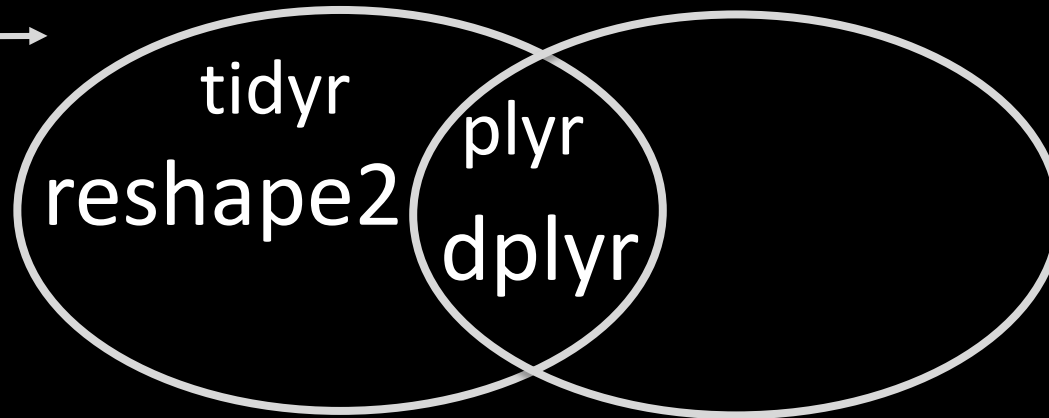
- Long format

- when, where, what, value



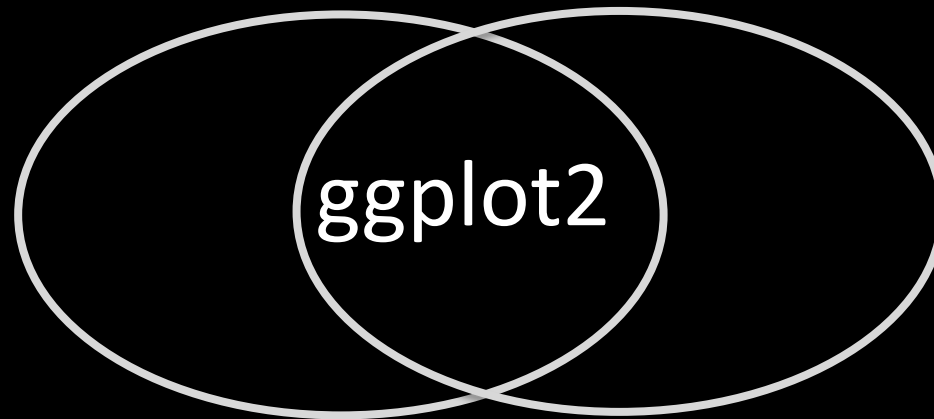
- Wide format

- lots of additional information



# Visualization: Exploratory Analysis

- ggplot2
  - geom\_point, geom\_line
  - transformation: linear scale
- ggplot2
  - geom\_tile, geom\_density
  - transformations: log





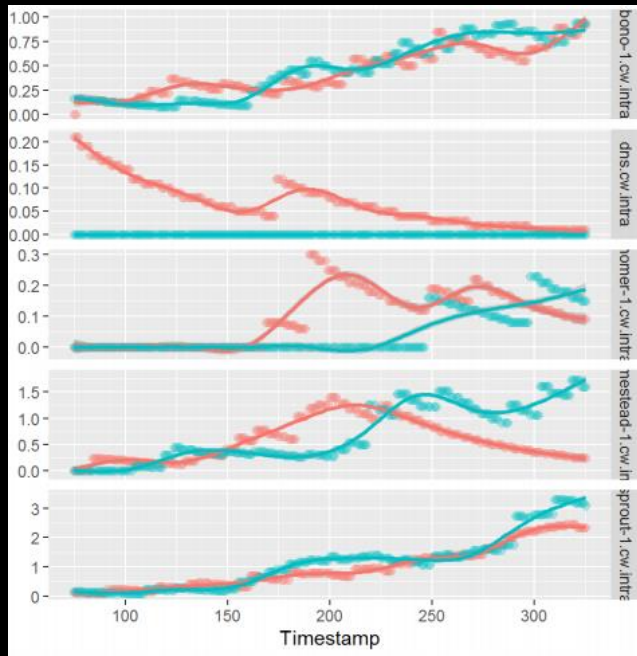
# Visualization: Exploratory Analysis

- ggplot2

- geom\_point, geom\_line
- transformation: linear scale

- ggplot2

- geom\_tile, geom\_density
- transformations: log



ggplot2

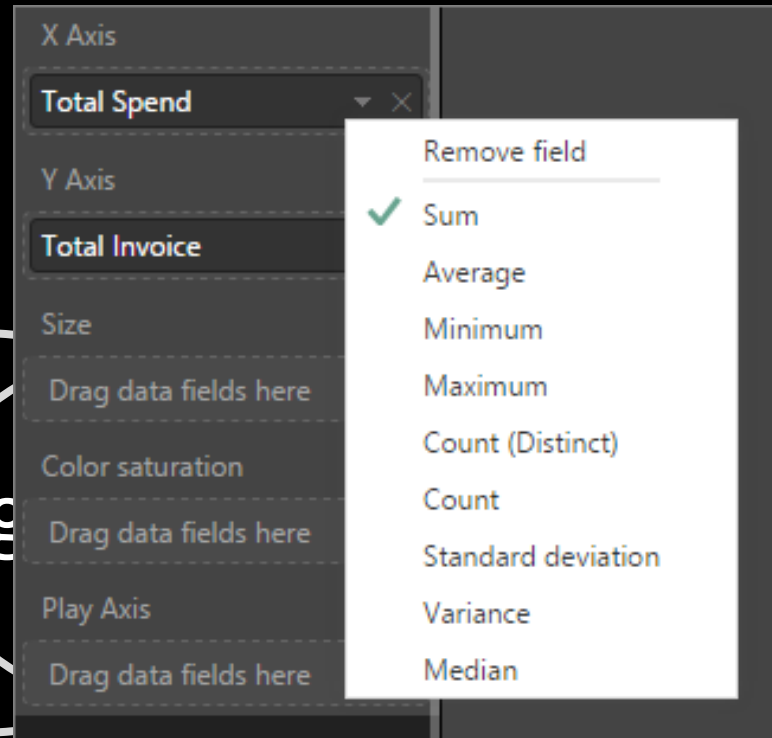
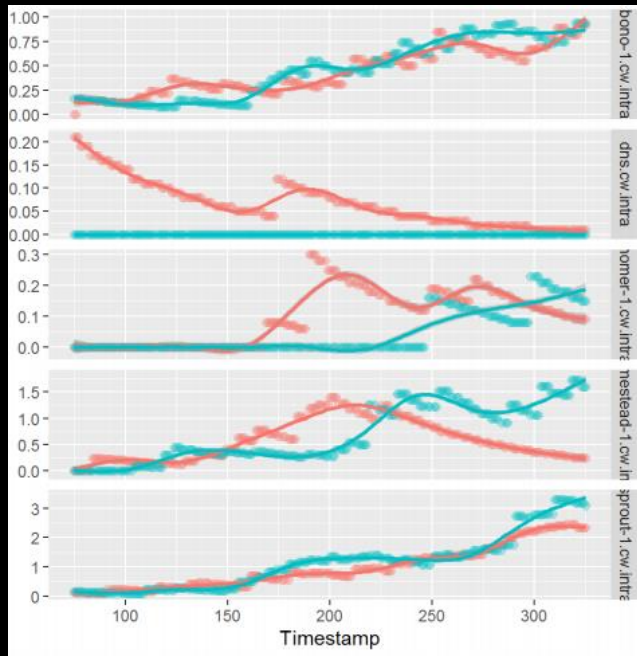
# Visualization: Exploratory Analysis

## ■ ggplot2

- geom\_point, geom\_line
- transformation: linear scale

## ■ ggplot2

- geom\_tile, geom\_density
- transformations: log



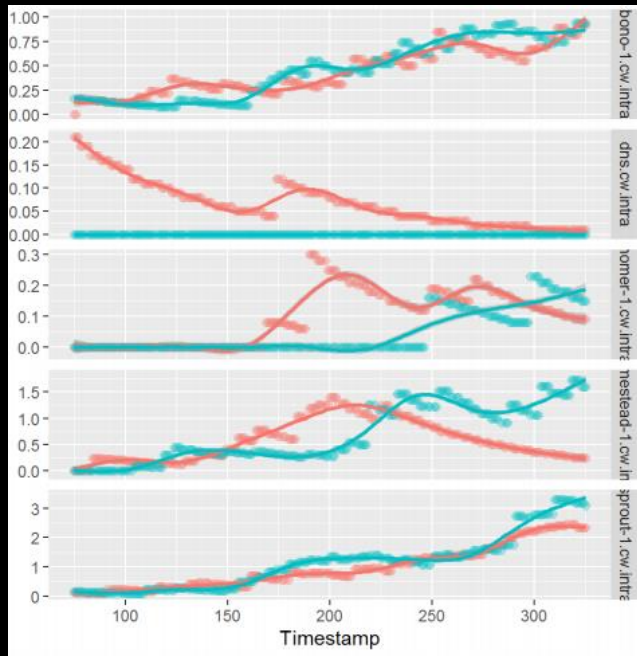
# Visualization: Exploratory Analysis

- ggplot2

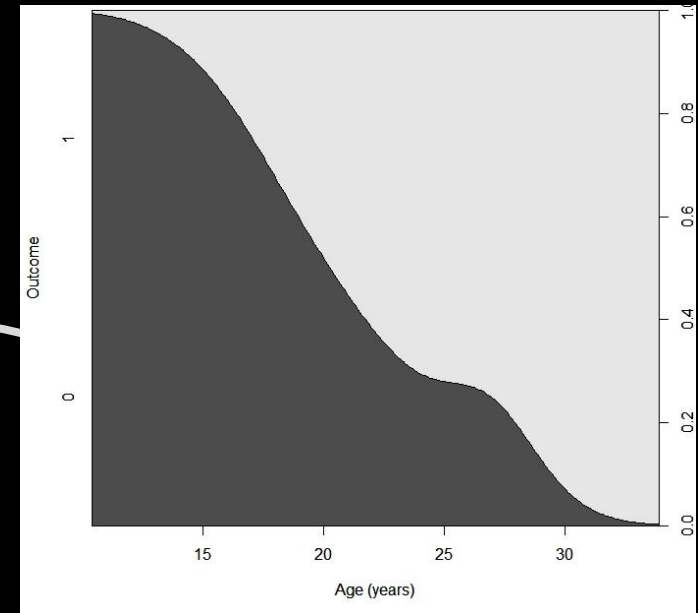
- geom\_point, geom\_line
- transformation: linear scale

- ggplot2

- geom\_tile, geom\_density
- transformations: log

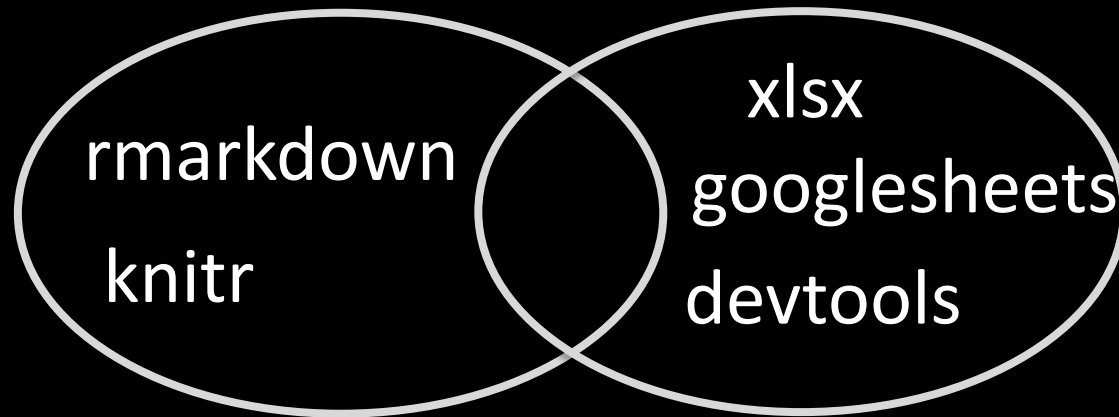


ggplot2

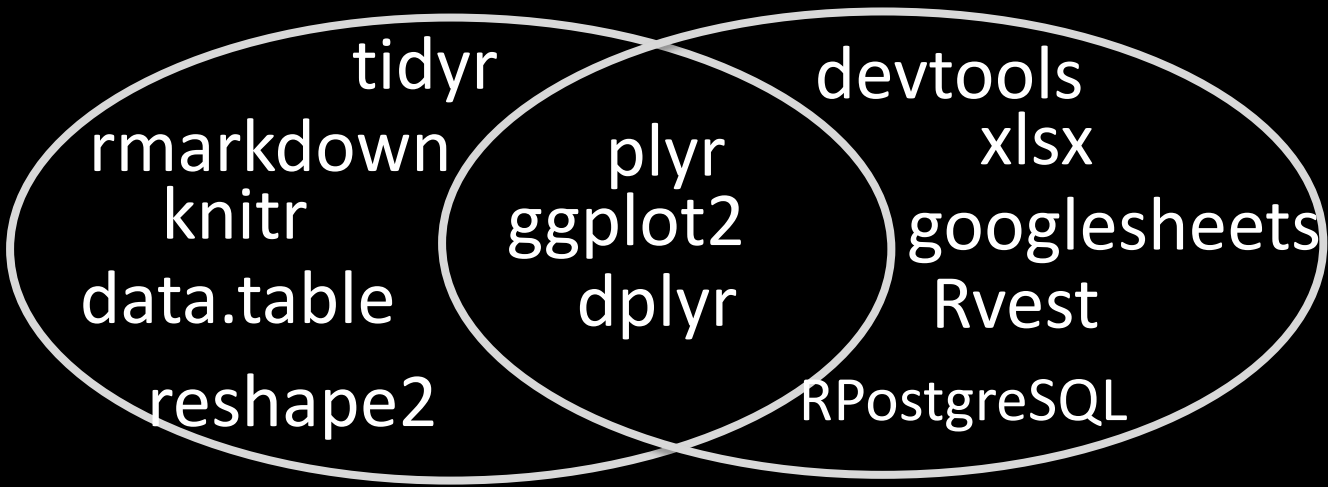


# Reporting/Collaboration

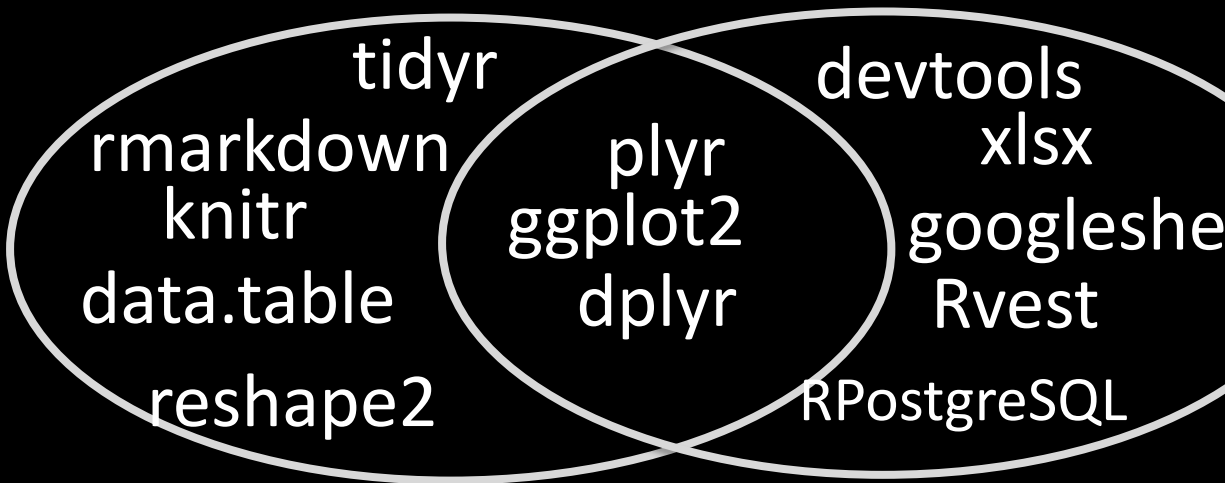
- Static pdf reports
- Excel/Google spreadsheets
- Package development



# Intersection



# Intersection



**Julia Silge** @juliasilge · Aug 16

ggplot2  
dplyr  
purrr  
devtools  
rmarkdown  
knitr  
shiny

[#7FavPackages](#) [#rstats](#)



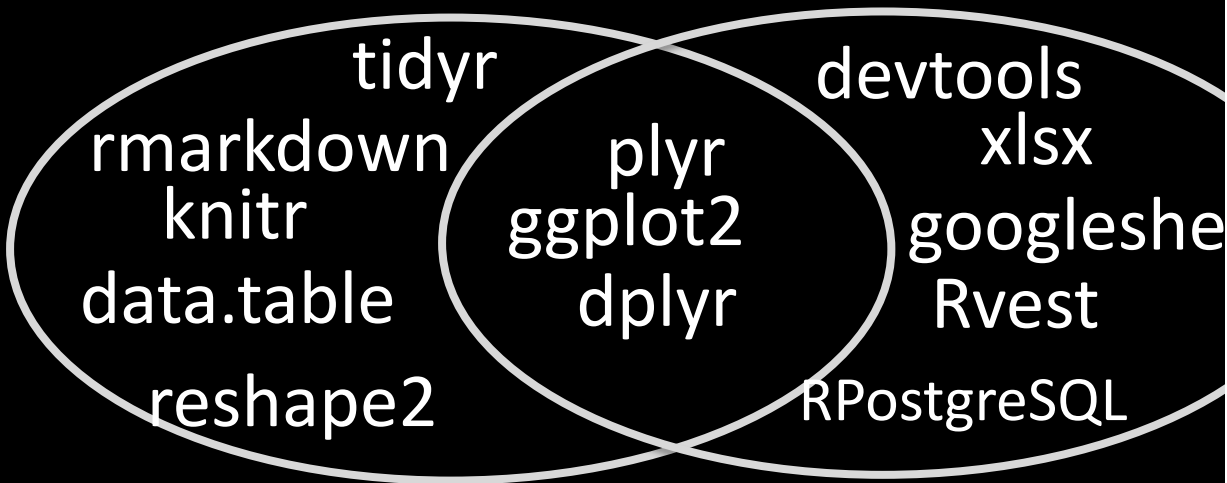
**Thomas Leeper** @thosjleeper · Aug 16

base  
stats  
tools  
utils  
grDevices  
graphics  
datasets

[#7FavPackages](#) [#rstats](#)



# Intersection



**Julia Silge** @julasilge · Aug 16

ggplot2  
dplyr  
purrr  
devtools  
rmarkdown  
knitr  
shiny

**#7FavPackages** #rstats



**Thomas Leeper** @thosjleeper · Aug 16

base  
stats  
tools  
utils  
grDevices  
graphics  
datasets

**#7FavPackages** #rstats



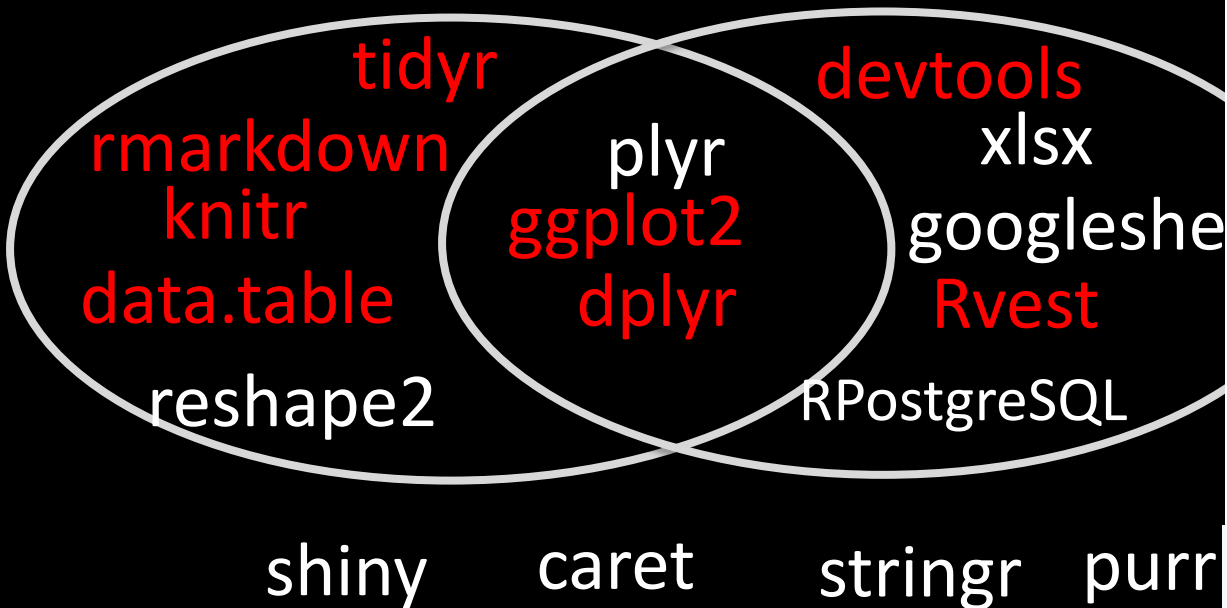
**Gabor Csardi** @GaborCsardi · Aug 16

devtools  
testthat  
jsonlite  
lintr  
shiny  
R6  
htmlwidgets

**#7FavPackages** #rstats



# Intersection



**Julia Silge** @julasilge · Aug 16

ggplot2  
dplyr  
purrr  
devtools  
rmarkdown  
knitr  
shiny

[#7FavPackages](#) [#rstats](#)



**Thomas Leeper** @thosjleeper · Aug 16

base  
stats  
tools  
utils  
grDevices  
graphics  
datasets

[#7FavPackages](#) [#rstats](#)



**Gabor Csardi** @GaborCsardi · Aug 16

devtools  
testthat  
jsonlite  
lintr  
shiny  
R6  
htmlwidgets

[#7FavPackages](#) [#rstats](#)





# Educational conclusions?

$$\begin{array}{l} \text{Confidence in data munging} \\ + \\ \text{Domain knowledge/curiosity} \\ \hline = \\ \text{Sense of achievement} \end{array}$$

