Shiny as a platform for collaborative data exploration

Imre Kocsis (kocsis@quanopt.com), research director, Quanopt Ltd.

Budapest BI Forum 2015, 2015.10.14.

About us

The company

- system/process analysis and optimization "consulting"
- automotive tooling development
- **R&D spinoff**
  - founders: professors & and lecturers
  - Dept. of Measurement and Inf. Systems @BME
  - course highlights: 'Big Data' Analysis Techniques, Cyber-Physical Systems, Intelligent Data Analysis, System Modelling
- Budapest R meetup: pizza, beer, presentations, 2 organizers

Data analysis competencies at Quanopt

Critical cloud applications

Manufacturing processes

Electronic equipment testing

Effort models for safety-critical software development processes

Fraud detection for service provider

Quanopt and R

You touch data - you have to speak some R (and you will see the light, we promise)

But a **manageable** investment - stats is the hard part, lang is easy

- Main use cases
  - **Visualization**: R is nigh unbeatable
  - "Exploratory Data Analysis"
  - Modeling - from descriptive models to neural networks
  - Data cleaning and tidying

Shiny: an overview

Shiny: a web framework for R

From RStudio

Web page layout and logic: write an R function

Server logic: write an R function

Page updates: reactive network of variables (as in Excel)

Run locally (RStudio!) / on shiny server / http://shinyapps.io

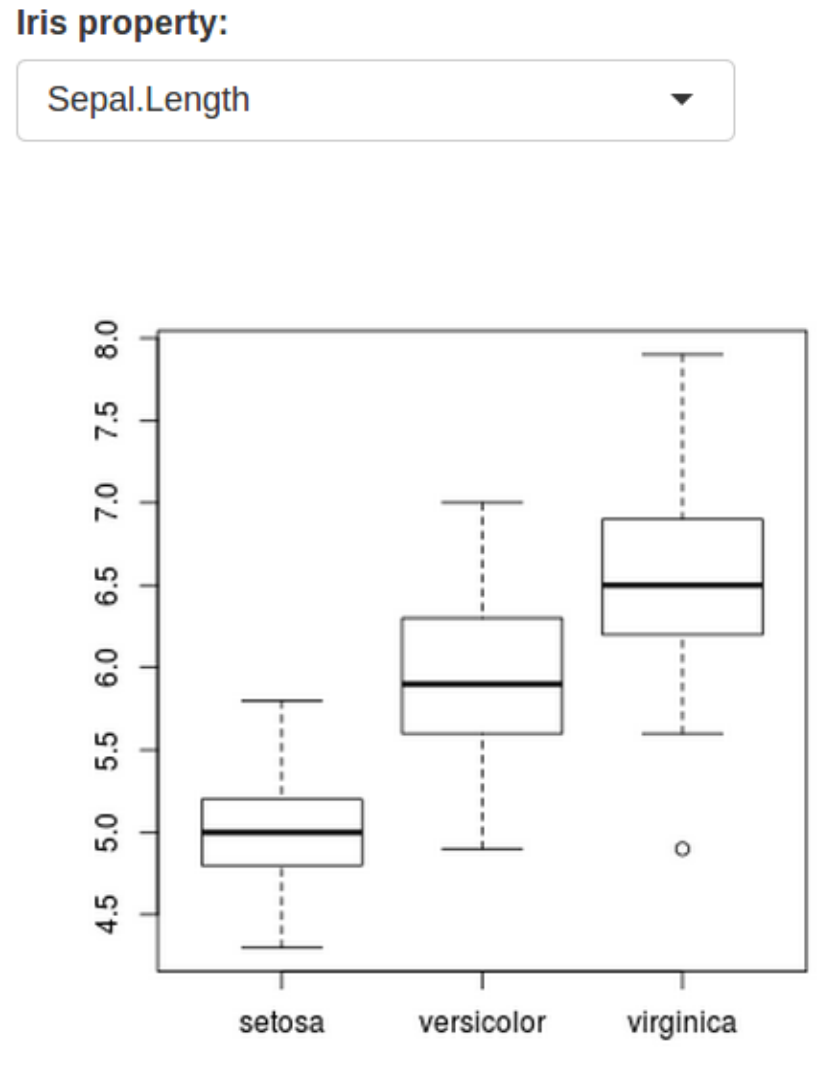
A simplistic example

ui.R

```
shinyUI(fluidPage(
  selectInput("iris_mvar",
    "Iris property:",
    choices=colnames(iris)[1:4]),
    plotOutput("boxplot")))
```

server.R

```
shinyServer(
  function(input,output){
    output$boxplot <- renderPlot(
      boxplot(get(input$iris_mvar) ~
      Species, data=iris))}))
```

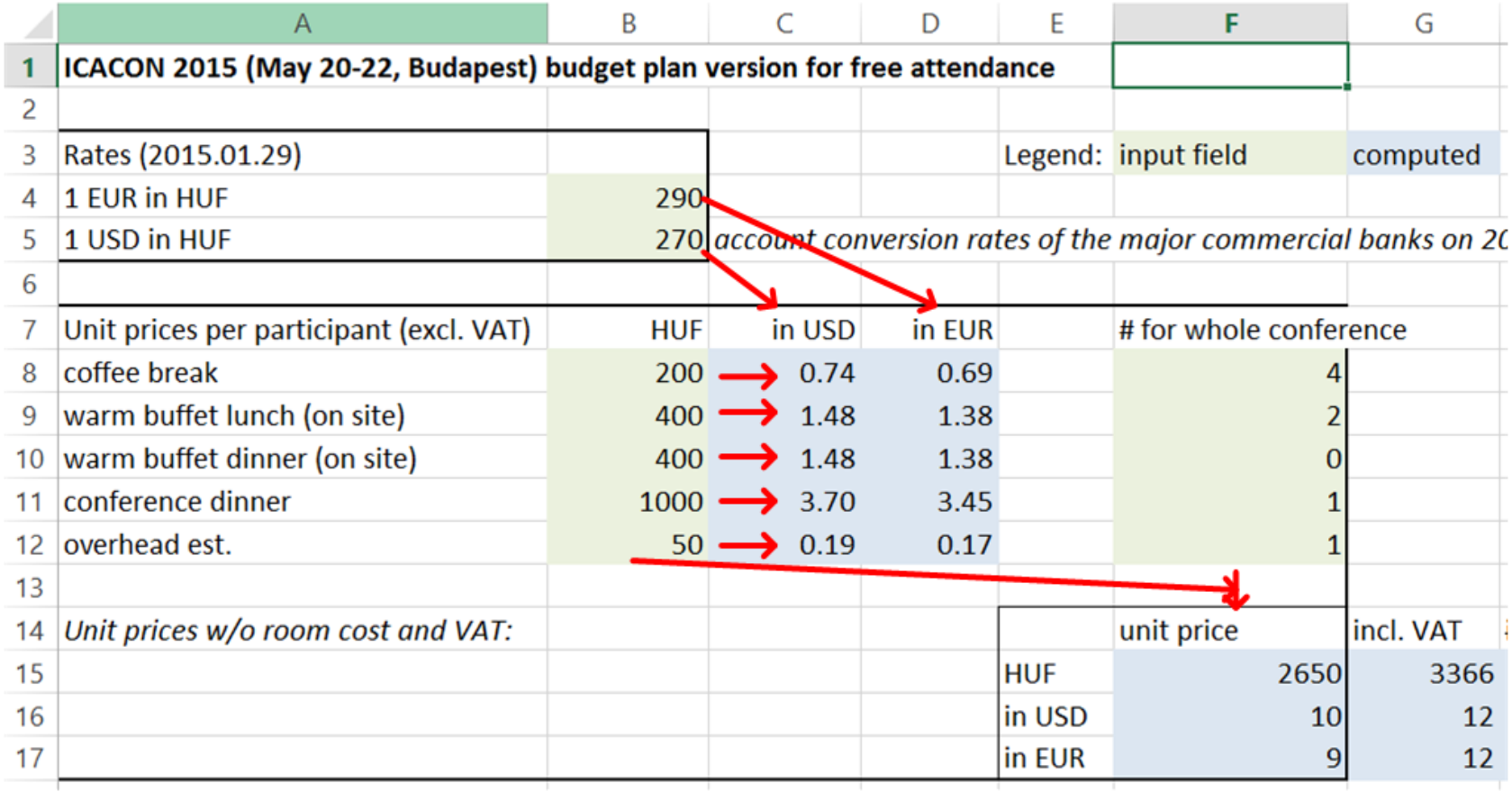


Reactive programming

	A	B	C	D	E	F	G
1	**ICACON 2015 (May 20-22, Budapest) budget plan version for free attendance**						
2							
3	Rates (2015.01.29)				Legend:	input field	computed
4	1 EUR in HUF	290					
5	1 USD in HUF	270	*account conversion rates of the major commercial banks on 2(*				
6							
7	Unit prices per participant (excl. VAT)	HUF	in USD	in EUR		# for whole conference	
8	coffee break	200	0.74	0.69		4	
9	warm buffet lunch (on site)	400	1.48	1.38		2	
10	warm buffet dinner (on site)	400	1.48	1.38		0	
11	conference dinner	1000	3.70	3.45		1	
12	overhead est.	50	0.19	0.17		1	
13							
14	*Unit prices w/o room cost and VAT:*					unit price	incl. VAT
15					HUF	2650	3366
16					in USD	10	12
17					in EUR	9	12

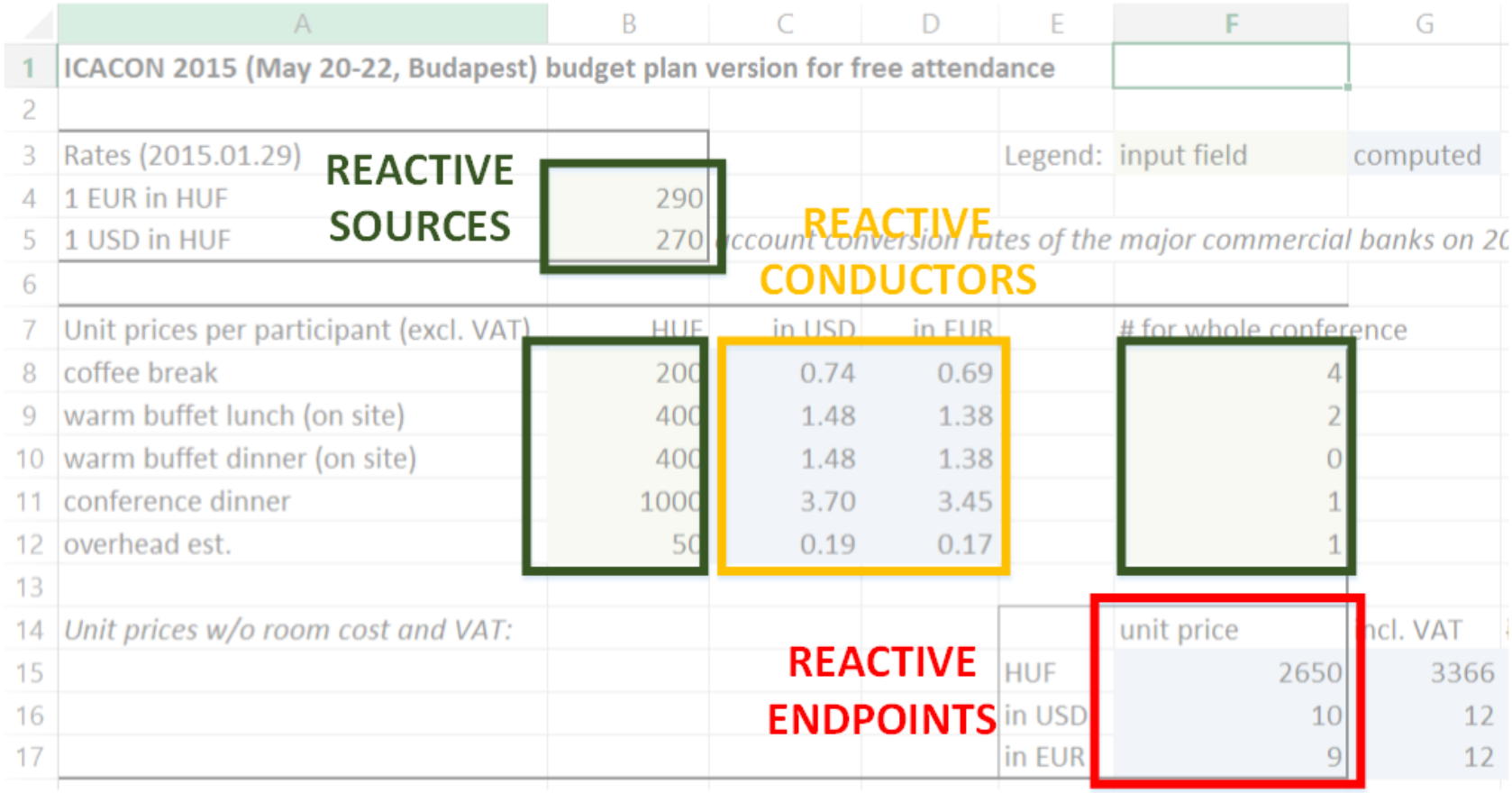
Note: prices have been obfuscated

Reactive programming



	A	B	C	D	E	F	G
1	**ICACON 2015 (May 20-22, Budapest) budget plan version for free attendance**						
2							
3	Rates (2015.01.29)				Legend:	input field	computed
4	1 EUR in HUF	290					
5	1 USD in HUF	270	*account conversion rates of the major commercial banks on 2(*				
6							
7	Unit prices per participant (excl. VAT)	HUF	in USD	in EUR		# for whole conference	
8	coffee break	200	0.74	0.69		4	
9	warm buffet lunch (on site)	400	1.48	1.38		2	
10	warm buffet dinner (on site)	400	1.48	1.38		0	
11	conference dinner	1000	3.70	3.45		1	
12	overhead est.	50	0.19	0.17		1	
13							
14	*Unit prices w/o room cost and VAT:*					unit price	incl. VAT
15					HUF	2650	3366
16					in USD	10	12
17					in EUR	9	12

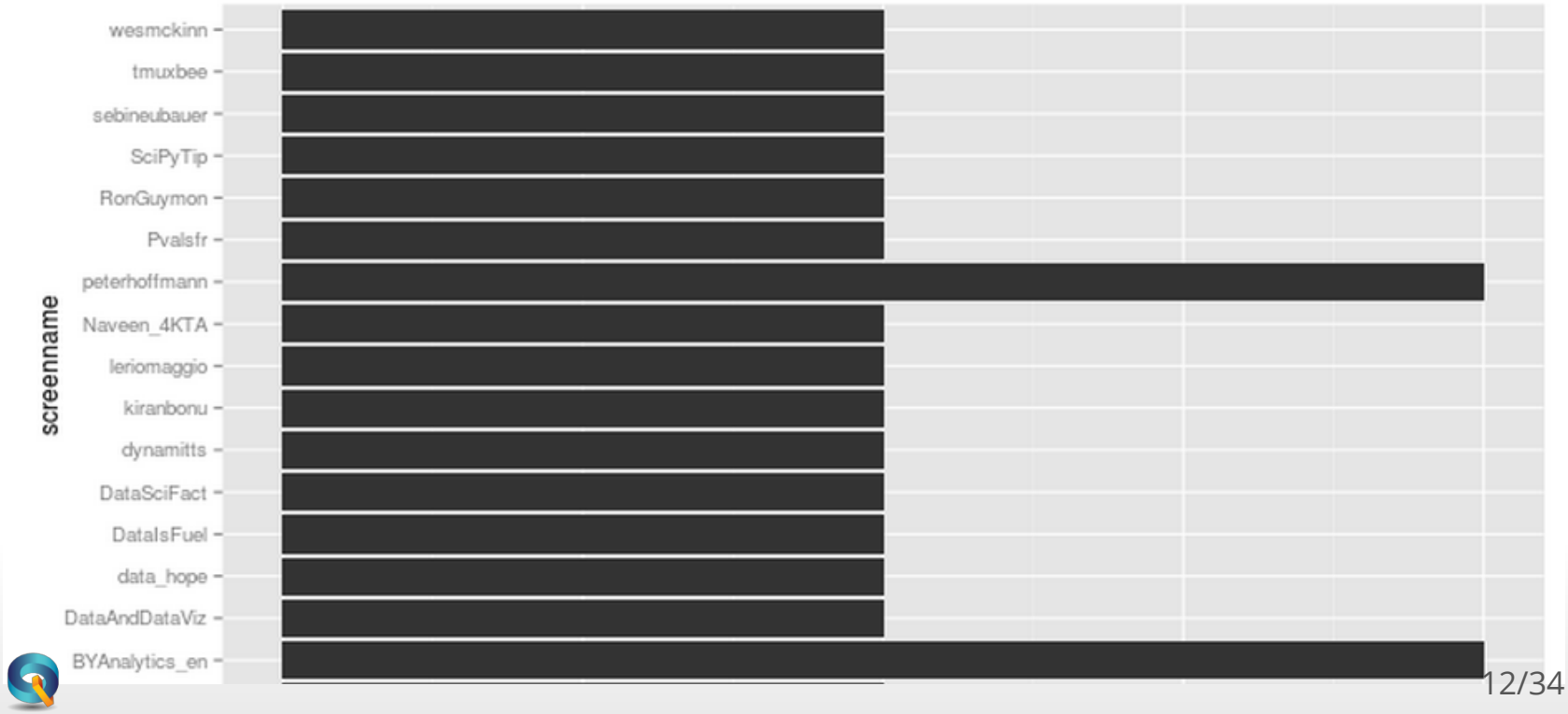
Reactive programming



Tweets mentioning "BudapestBI" - query with button

Go!

Rotate figure



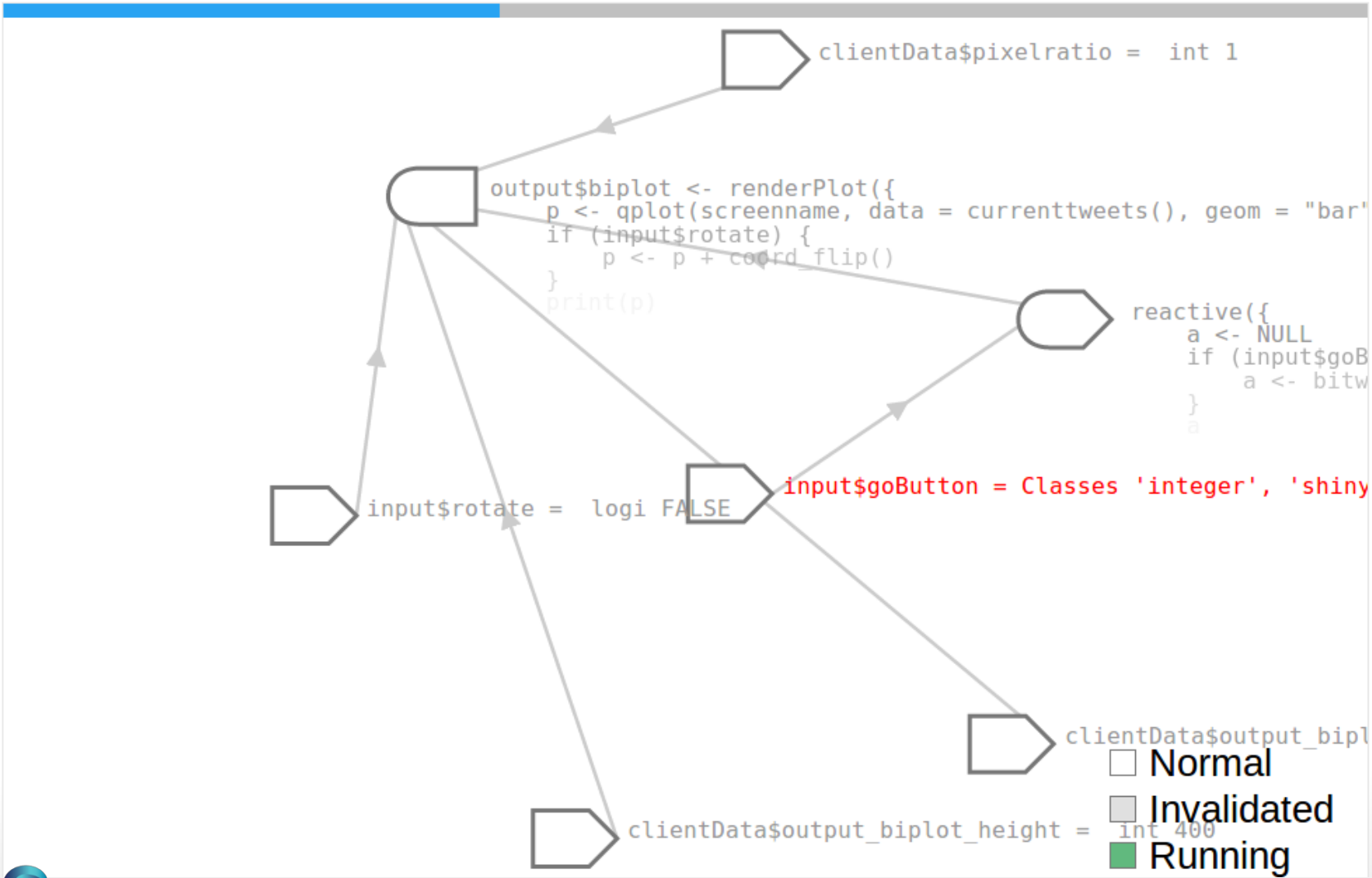
The code - server

```
function(input, output){
    currenttweets <- reactive({
      a <- NULL
      if(input$goButton > 0) {
        a <- bitweeters()}
      a
    })

    output$biplot <- renderPlot({
      p <- qplot(screenname, data=currenttweets(), geom="bar")
      if(input$rotate) {p <- p + coord_flip()}
      print(p)
    })
```

And for the sake of completeness

```
library(twitteR)
library(ggplot2)

bitweeters <- function(){
  origop <- options("httr_oauth_cache")
  options(httr_oauth_cache=TRUE)
  setup_twitter_oauth(conskey,conssecr,actok, acsecr)
  options(httr_oauth_cache=origop)

  tweets <- searchTwitter("BudapestBI", n = 100)

  scrn <- sapply(tweets, screenName)
  crtd <- sapply(tweets, created)

  data.frame(screenname=scrn, created=crtd)
}
```

Invalidation and recomputation

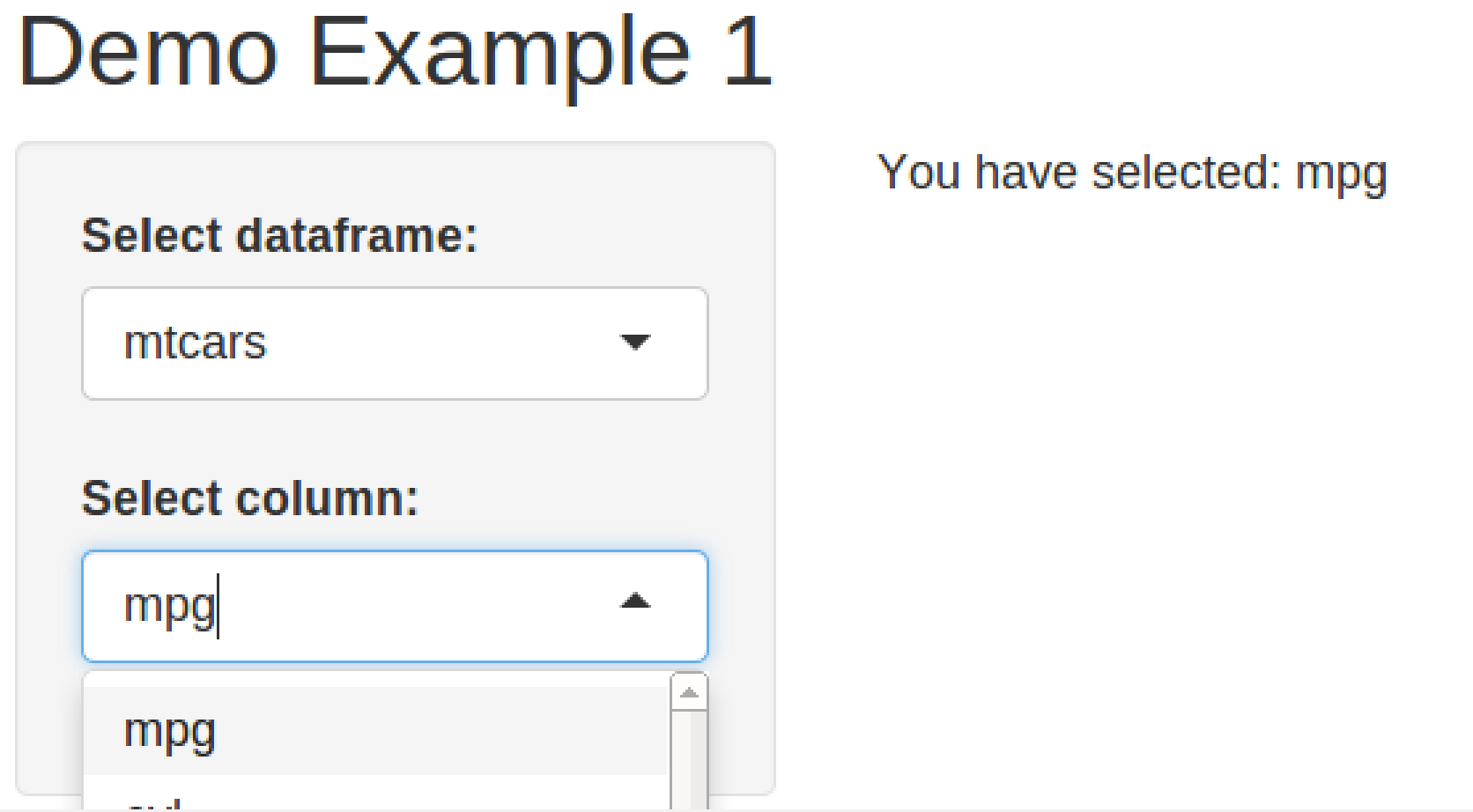


Other technicalities

- you can isolate the usage of a reactive var
- observers for side effects upon changes
- all the usual HTML inputs
- reasonable layout control
- reasonable set of outputs (with renderers) + htmlwidgets
- UI customization: HTML, JavaScript, CSS
- server function can have session parameter
- **important, but not covered: interactive graphics**

Dynamic UI

or towards intractable hacking



renderUI for uiOutput

```
function(input, output) {
  output$colselect <- renderUI({
    if(input$dataSelect == "cars") {
      selectInput("colSelect", "Select column:", names(dataframe1))
    } else {
      selectInput("colSelect", "Select column:", names(dataframe2))
    }
  })

  output$selected <- renderText({
    paste("You have selected: ", input$colSelect)
  })}
```

Could be different control types

And you can do cascading (**but don't**!)

Use case #1

EDA on public cloud benchmark campaigns

Scenario

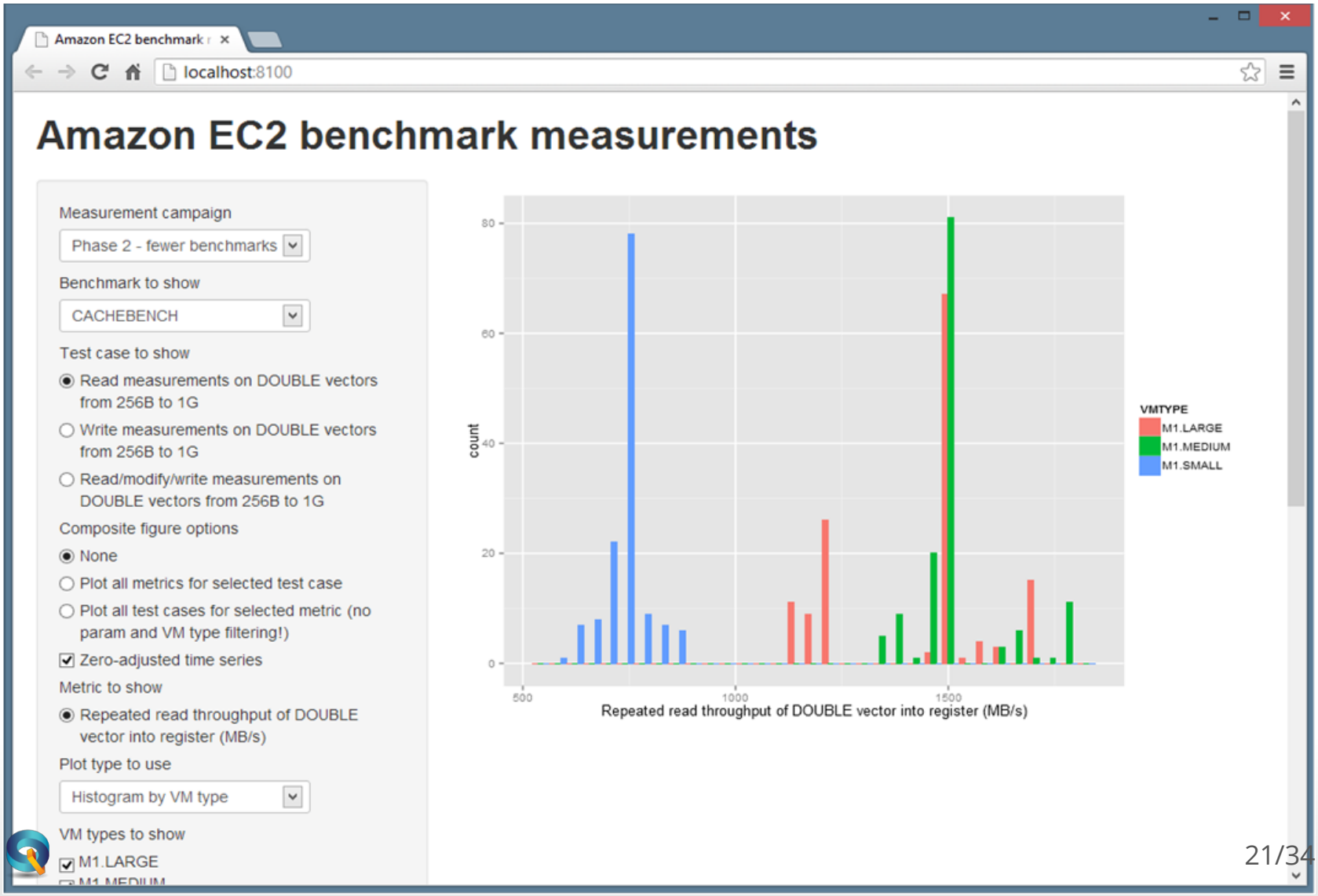
Cloud automation & availability project

An awful lot of perf. benchmark campaigns

To specify "the plots" for the Big BI DW Tool

R/Shiny begun as a band-aid

Results


Amazon EC2 benchmark
localhost:8100
Amazon EC2 benchmark measurements
Measurement campaign
Phase 2 - fewer benchmarks
Benchmark to show
CACHEBENCH
Test case to show
Read measurements on DOUBLE vectors from 256B to 1G
Write measurements on DOUBLE vectors from 256B to 1G
Read/modify/write measurements on DOUBLE vectors from 256B to 1G
Composite figure options
None
Plot all metrics for selected test case
Plot all test cases for selected metric (no param and VM type filtering!)
Zero-adjusted time series
Metric to show
Repeated read throughput of DOUBLE vector into register (MB/s)
Plot type to use
Histogram by VM type
VM types to show
M1.LARGE
M1.MEDIUM
count
0
20
40
60
80
500
1000
1500
Repeated read throughput of DOUBLE vector into register (MB/s)
VMTYPE
M1.LARGE
M1.MEDIUM
M1.SMALL

Lessons learned

Very positive reception (Because shiny.)

In the given context: visualization >> "statistics"

First usable prototypes: very fast

Put it in a VPN - and suddenly it's a tool for **collaboration**

DB normalization and evolution massively worked against us

Deep, dynamic tree of DB-driven input controls: **AVOID!**

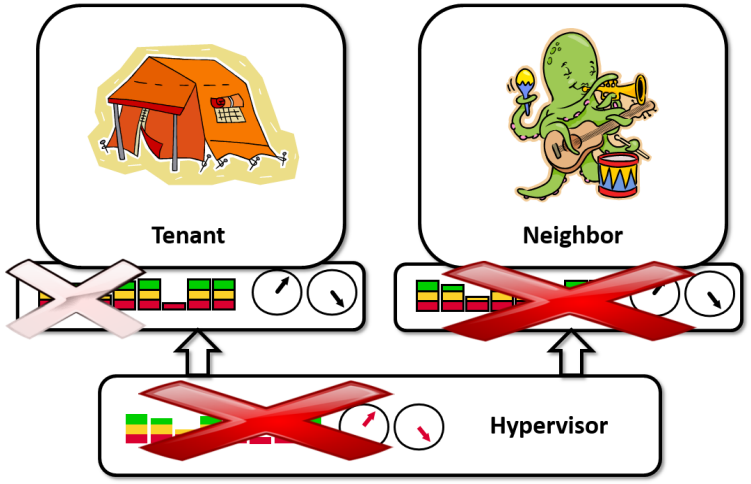
The whole thing got a bit sluggish (mostly ggplot2)

If it works… (?)

Similar projects, similar usage

Electronic equipment testing (BIG data)

Currently: fault injection on telco cloud applications



Web-based, "mouse-only" exploration

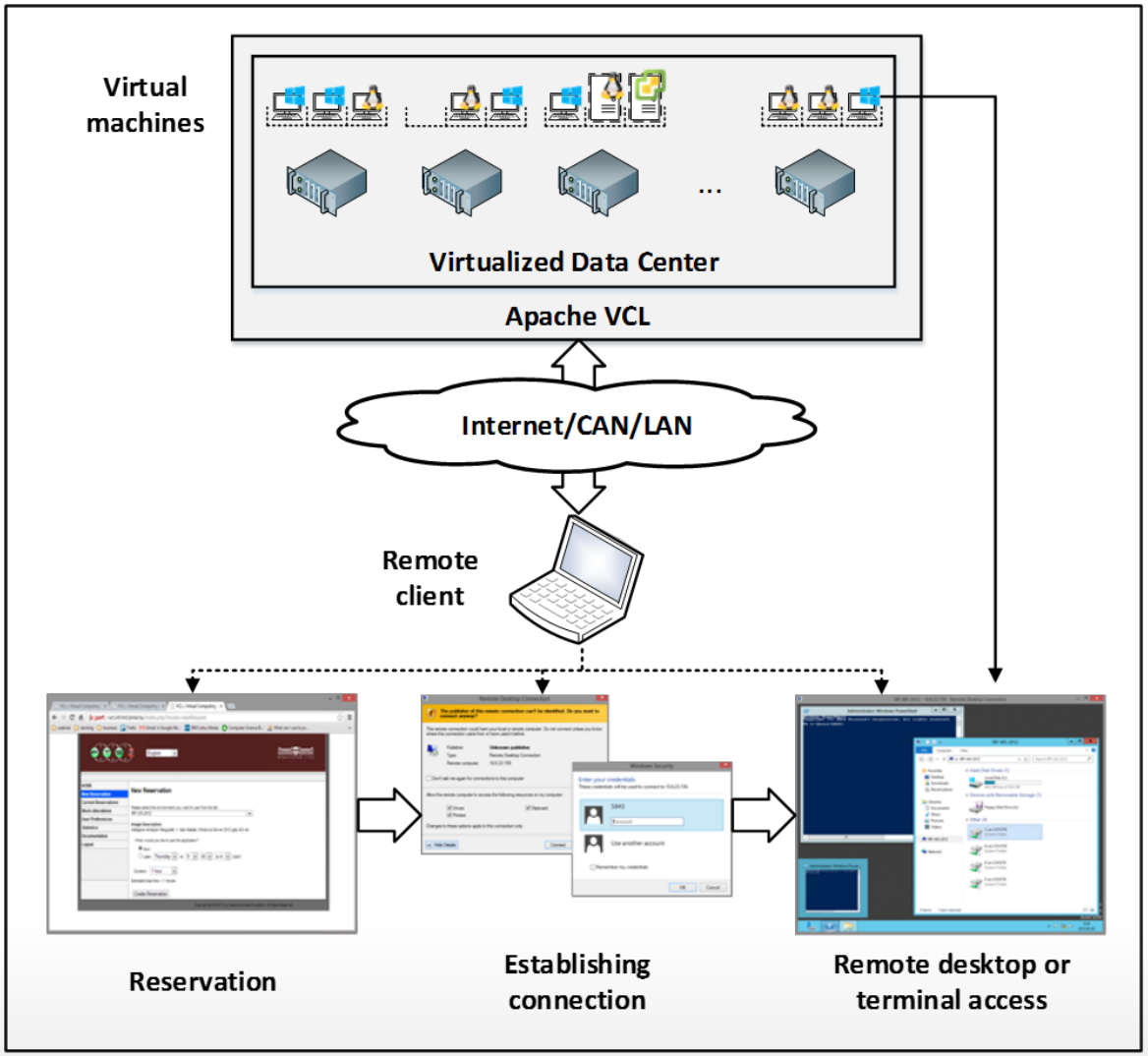
Small teams, project-internal usage; findings recorded separately

Source of all griefs: ID variables

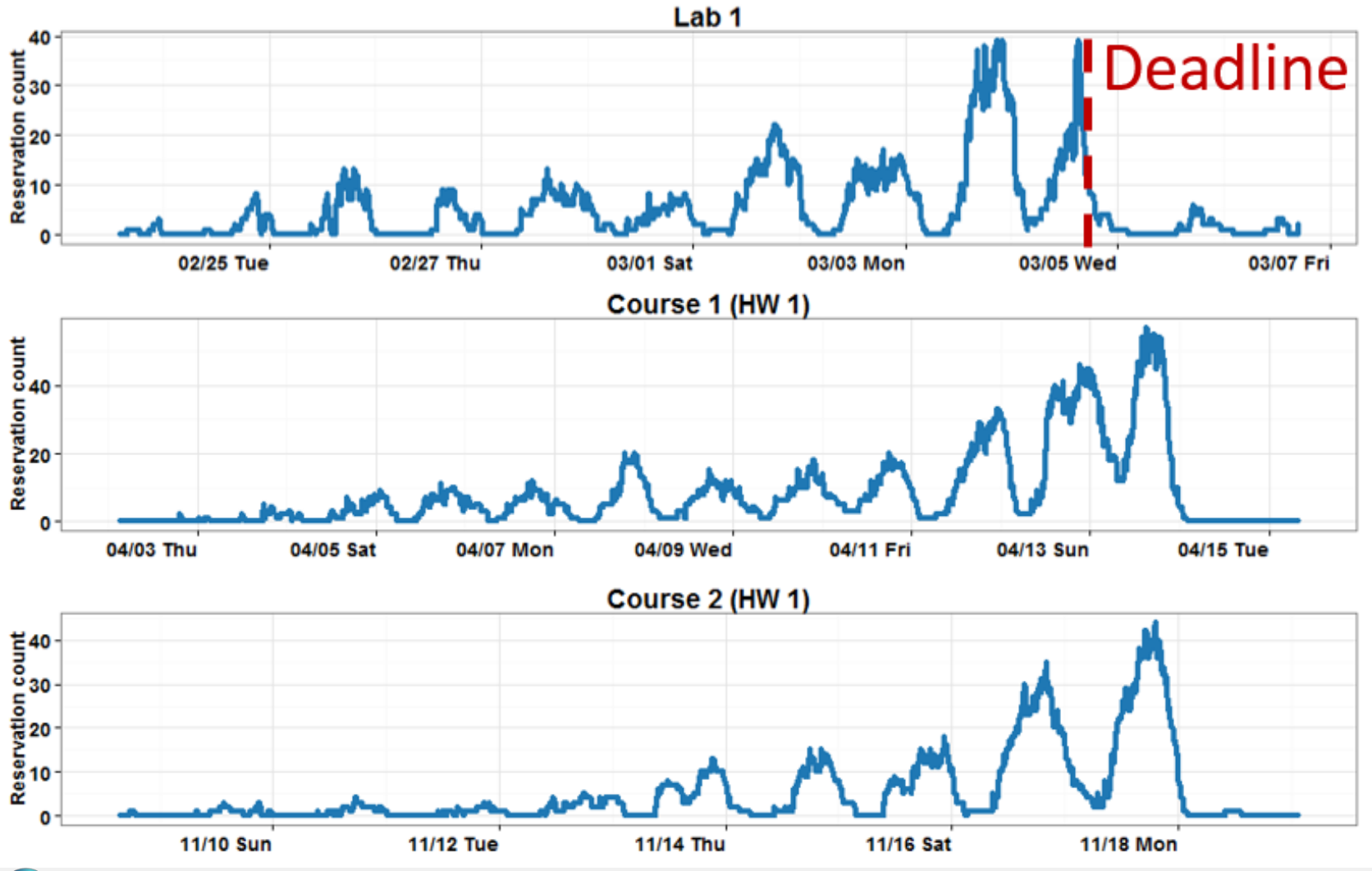
"Decision making support"

Let stakeholders convince themselves

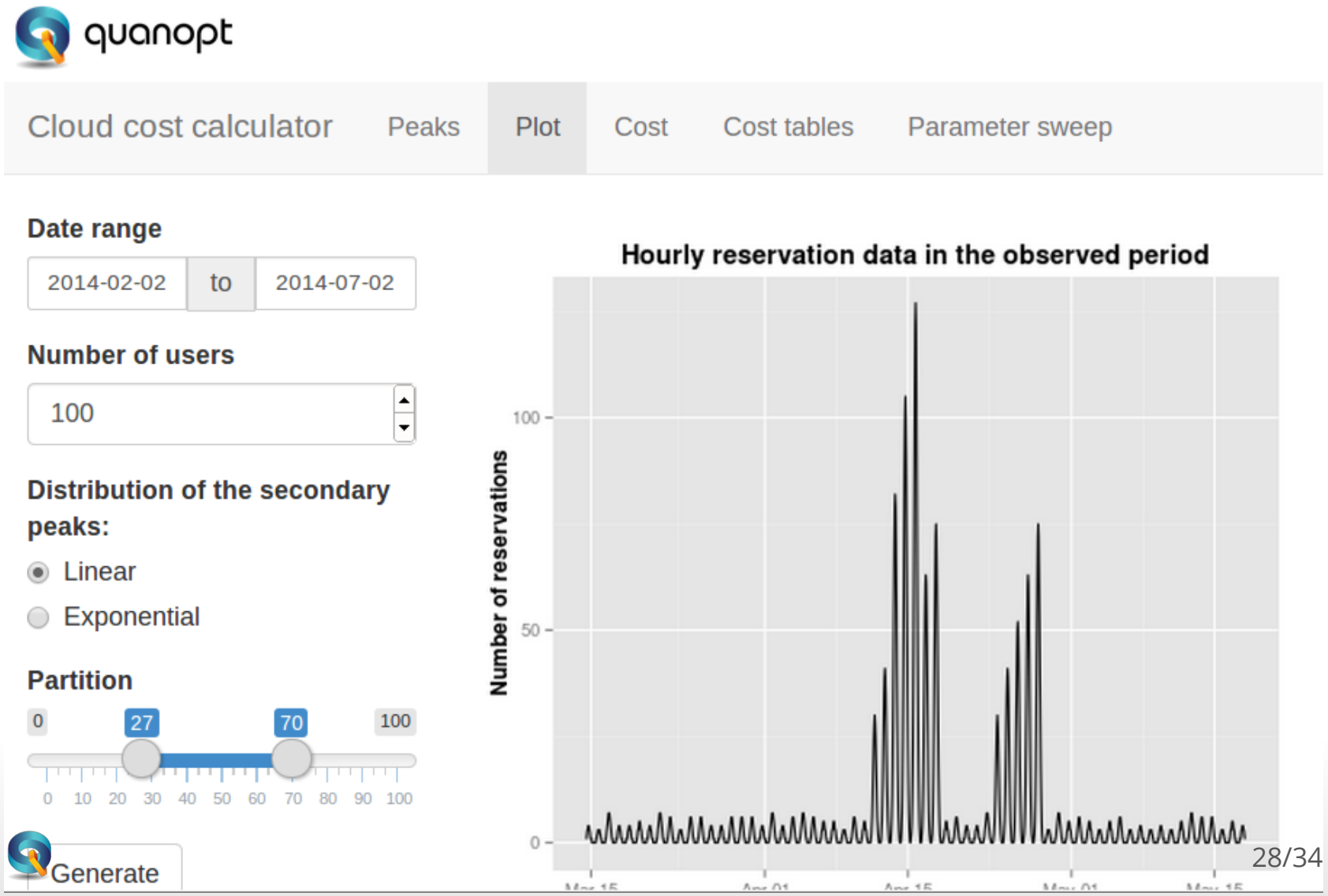
Educational cloud: Apache VCL



How do students behave?



The app


quanopt
Cloud cost calculator
Peaks
Plot
Cost
Cost tables
Parameter sweep
Date range
2014-02-02
to
2014-07-02
Number of users
100
Distribution of the secondary peaks:
Linear
Exponential
Partition
0
27
70
100
0 10 20 30 40 50 60 70 80 90 100
Generate
Hourly reservation data in the observed period
Number of reservations
100
50
0

My impression of shiny

Shiny: pro's

- Inside its design envelope: a breeze
- For me this includes no JavaScript
- Almost full R ecosystem
- By far not only "toy apps"
- … But you probably want to stop at "workgroup applications"

Shiny: con's and things to keep in mind

- No notion of
  - page flow (single page)
  - analysis flow
  - recording & replaying app state
- Keep layout dynamics unsophisticated
- Is the reactive model really enough for you?
- You still develop R (with "reactivity") - maintenance?

Shiny: con's and things to keep in mind

- Stepping out of the box: R scripting morphs into web development
- Don't confuse it with full-fledged tools (no Watson for free)
- optimal case: your code and data in R; OR you need something from R
- This can be simply ggplot2, mind you
- Server/PaaS pricing
- (Interactive graphics - incl. linked selection & highlighting)

Final thoughts

- Fun! :)
- Check it out: http://shiny.rstudio.com
- Professional setting: think about the risks (and whys)
- Under development: making objects of the analyzed system first-class citizens in shiny-based visualization
- Currently: too observation-based (interpretation?)

Thanks

http://quanopt.com

Dynamic presentation:
https://github.com/imrekocsis/biforum2015

Mail
kocsis@quanopt.com
ikocsis@mit.bme.hu

Special thanks from the team to
László Gönczy, Gergő Kincses, Ágnes Salánki, György Nádudvari