

VisRseq: R-based visual analytics software for sequencing data

Hamid Younesy

Graphics Usability and Visualization, Simon Fraser University, Canada

Torsten Möller

Visualization and Data Analysis Group, University of Vienna, Austria

Matthew C. Lorincz

Department of Medical Genetics, The University of British Columbia, Canada

Mohammad M. Karimi

Biomedical Research Centre, The University of British Columbia, Canada

Steven J.M. Jones

Genome Sciences Centre, BC Cancer Agency, Canada





Assignment

Assignment: Analyzing ML model parameterizations

- How does ML parameterization affects performance?
- Use the full (100,000 row) dataset for analysis
- Write a report discussing what you found
- Single PDF document
- Plots should be generated using only VisRseq
- Other software is ok to add text and arrange/annotate graphics

Before we begin...

- download visrseq

<http://bit.ly/2gufWsB>

- clone assignment repo

`git clone https://github.com/hyounesy/cass2017_vis.git`

<http://bit.ly/2gufWsB>

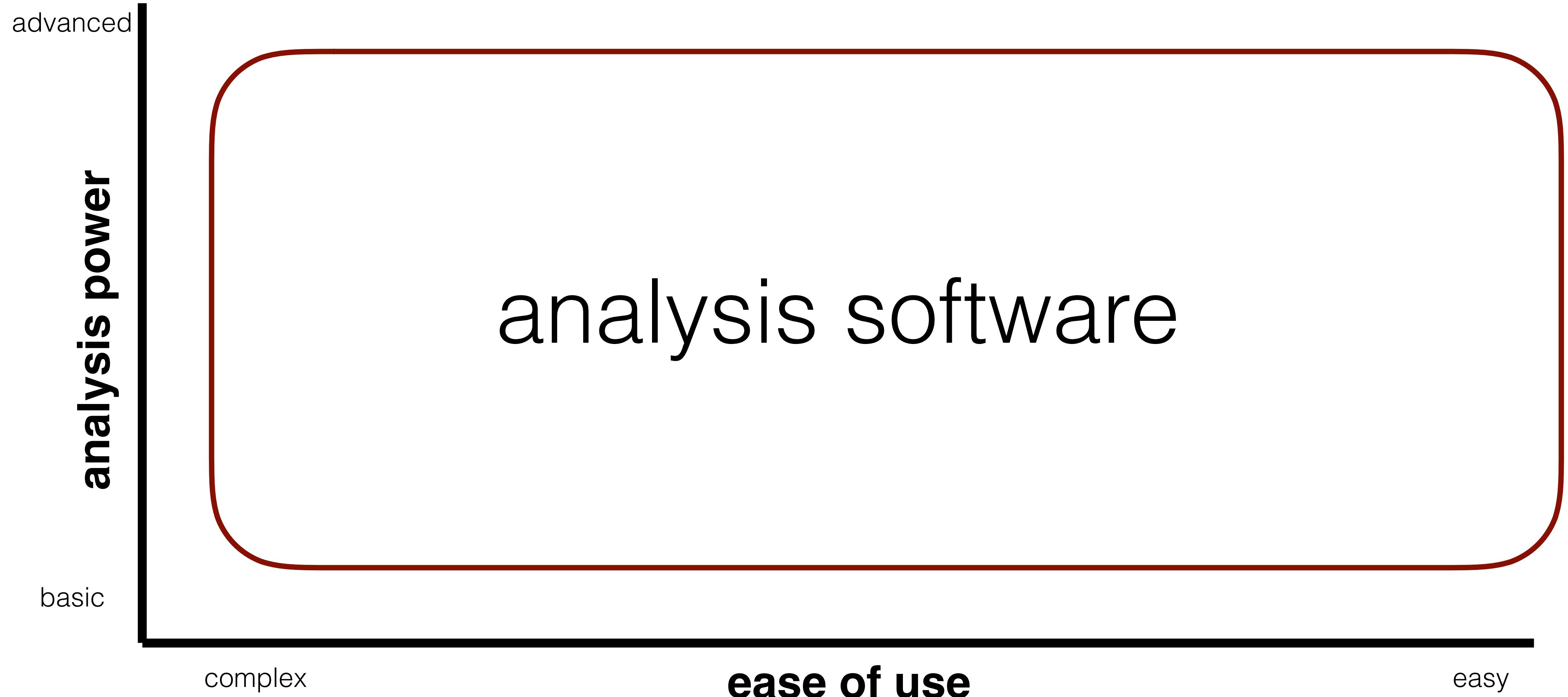
git clone https://github.com/hyounesy/cass2017_vis.git

motivation

motivation

<http://bit.ly/2gufWsB>

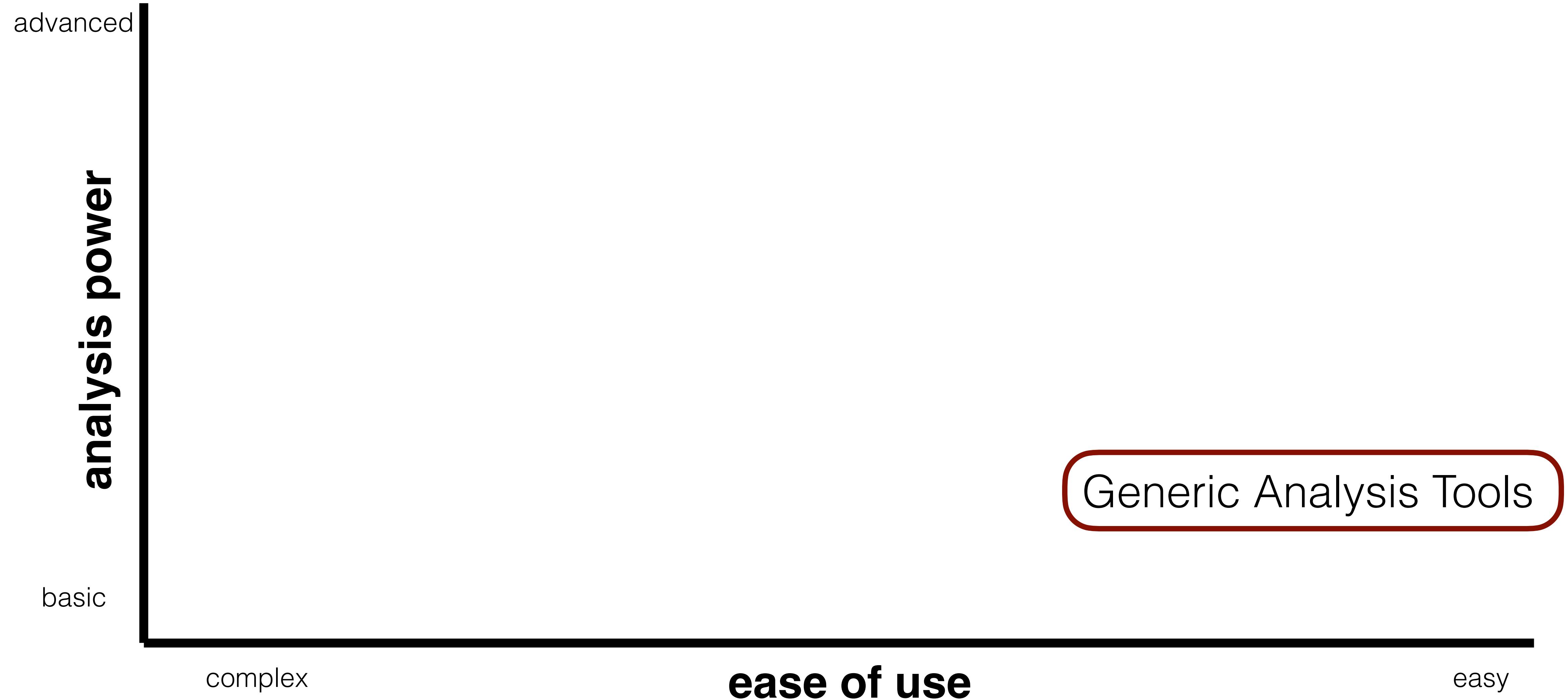
git clone https://github.com/hyounesy/cass2017_vis.git



motivation

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git



motivation

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git

advanced

analysis power

basic

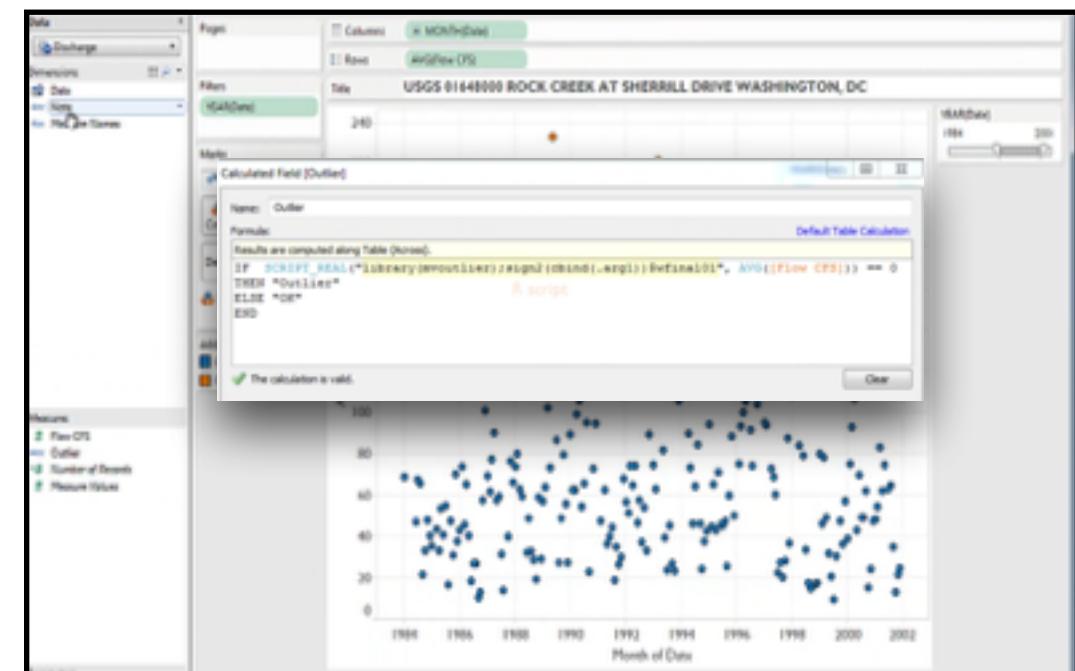
complex

ease of use

easy

Category	Term	Fold Enrichr	PValue	Count	%
GOTERM_BP_GO_0006412	7.16288483	2.85E-19	translation	33	
GOTERM_BP_GO_0034645	2.58363459	1.37E-09	cellular macromolecule biosynthetic proc	42	
GOTERM_BP_GO_0009059	2.56246481	1.77E-09	macromolecule biosynthetic process	42	
GOTERM_BP_GO_0010467	2.28422179	2.51E-07	gene expression	39	
GOTERM_BP_GO_0044267	2.09523711	3.58E-06	cellular protein metabolic process	38	
GOTERM_BP_GO_0019538	1.95086485	4.53E-06	protein metabolic process	42	
GOTERM_BP_GO_0044249	1.78235807	8.96E-06	cellular biosynthetic process	47	
GOTERM_BP_GO_0009058	1.70532395	3.13E-05	biosynthetic process	47	
GOTERM_BP_GO_0030199	29.7840112	3.04E-04	collagen fibril organization	4	
GOTERM_BP_GO_0030198	18.8111342	4.50E-04	extracellular matrix organization	5	
GOTERM_BP_GO_0022900	5.39395479	6.68E-04	electron transport chain	8	
GOTERM_BP_GO_0043062	8.15514593	0.00319551	extracellular structure organization	5	
GOTERM_BP_GO_0022904	12.0181098	0.00415644	respiratory electron transport chain	4	
GOTERM_BP_GO_0043170	1.33494519	0.00590397	macromolecule metabolic process	52	
GOTERM_BP_GO_0044237	1.22562383	0.00676271	cellular metabolic process	67	
GOTERM_BP_GO_0044260	1.36054073	0.00680266	cellular macromolecule metabolic proces	48	
GOTERM_BP_GO_0045333	4.82417083	0.00783823	cellular respiration	6	
GOTERM_BP_GO_0051216	19.7605459	0.00980216	cartilage development	3	
GOTERM_BP_GO_0008292	171.258065	0.0115197	acetylcholine biosynthetic process	2	
GOTERM_BP_GO_0009105	5.63348896	0.01164616	glycoprotein metabolic process	5	
GOTERM_BP_GO_0009987	1.1327528	0.01372666	cellular process	80	
GOTERM_BP_GO_0006091	2.59875667	0.01379198	generation of precursor metabolites and i	10	
GOTERM_BP_GO_0032501	1.70689101	0.01607076	multicellular organismal process	21	
GOTERM_BP_GO_0042136	85.6290323	0.02190813	neurotransmitter biosynthetic process	2	
GOTERM_BP_GO_0008291	85.6290323	0.02290813	acetylcholine metabolic process	2	

Excel



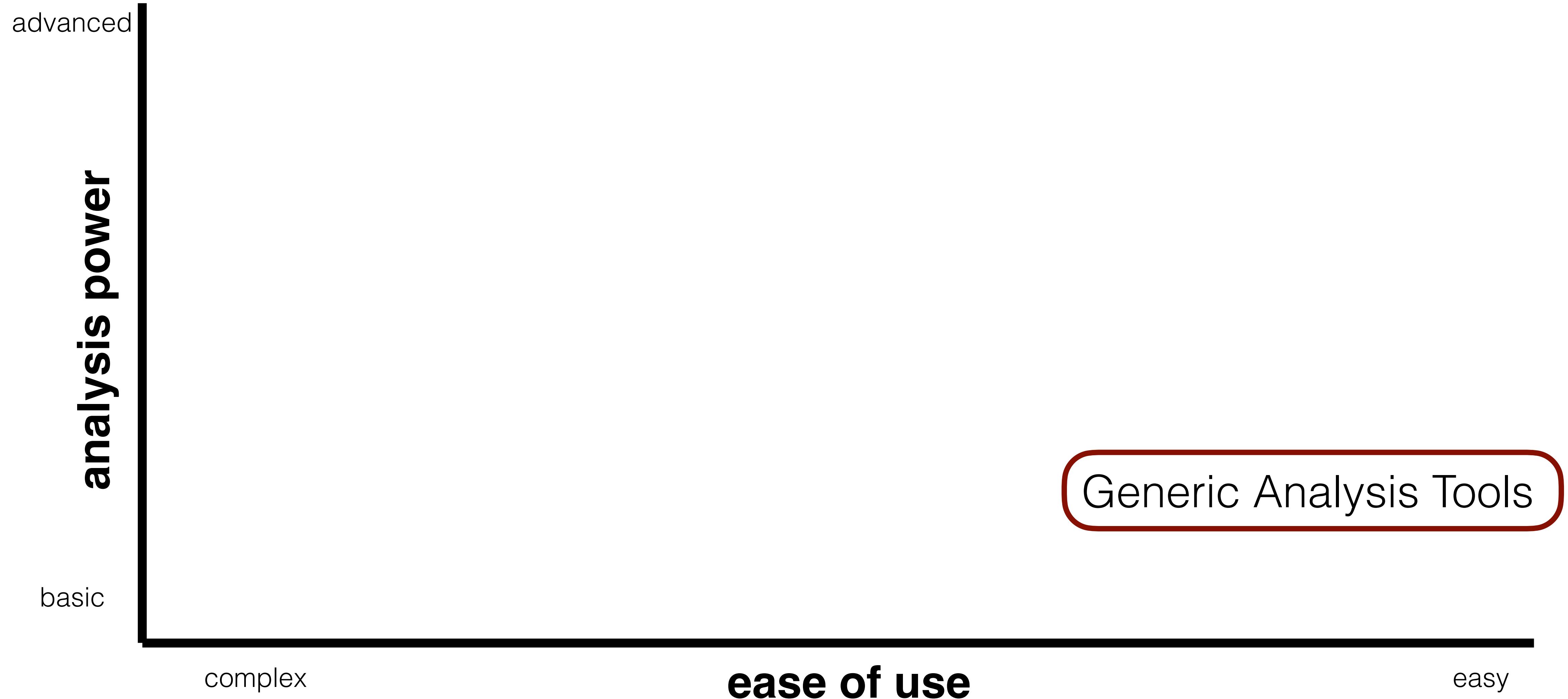
Tableau

Generic Analysis Tools

motivation

<http://bit.ly/2gufWsB>

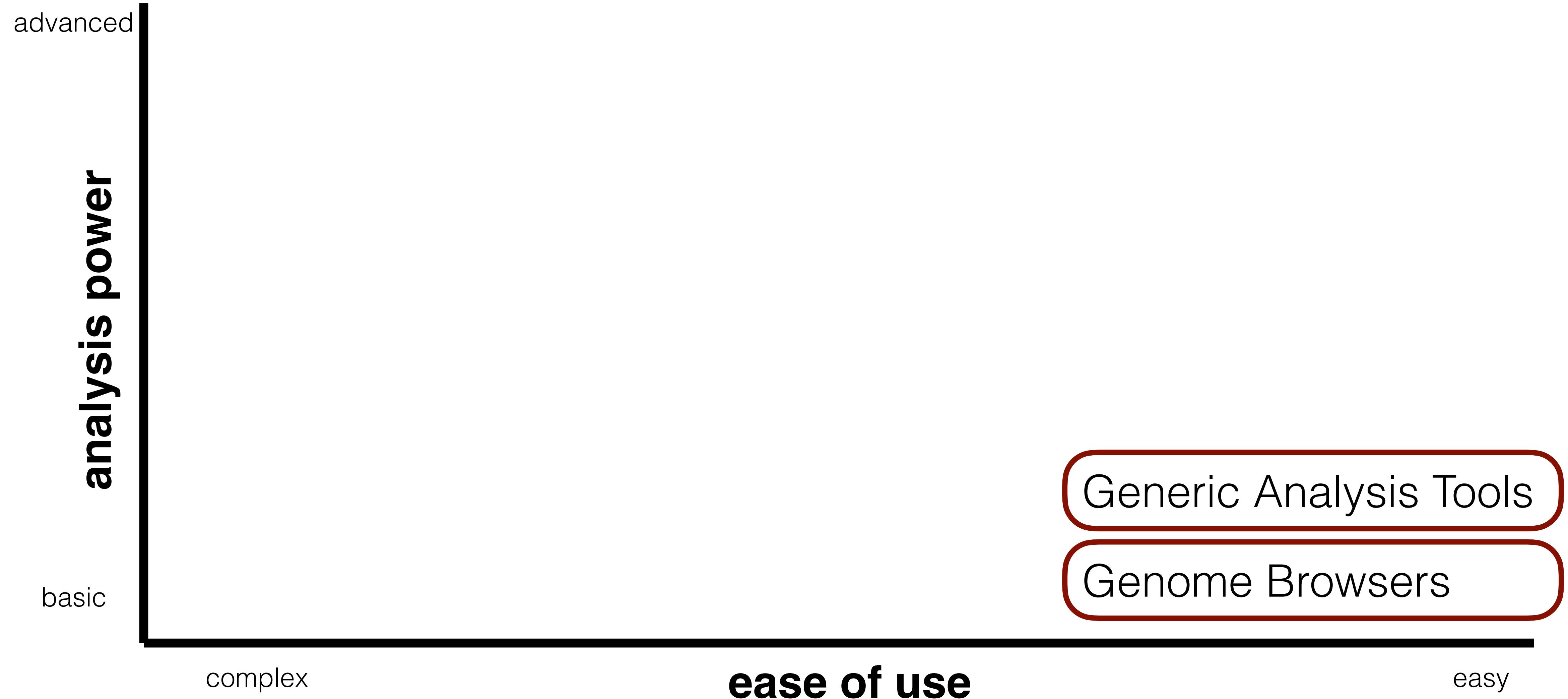
git clone https://github.com/hyounesy/cass2017_vis.git



motivation

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git

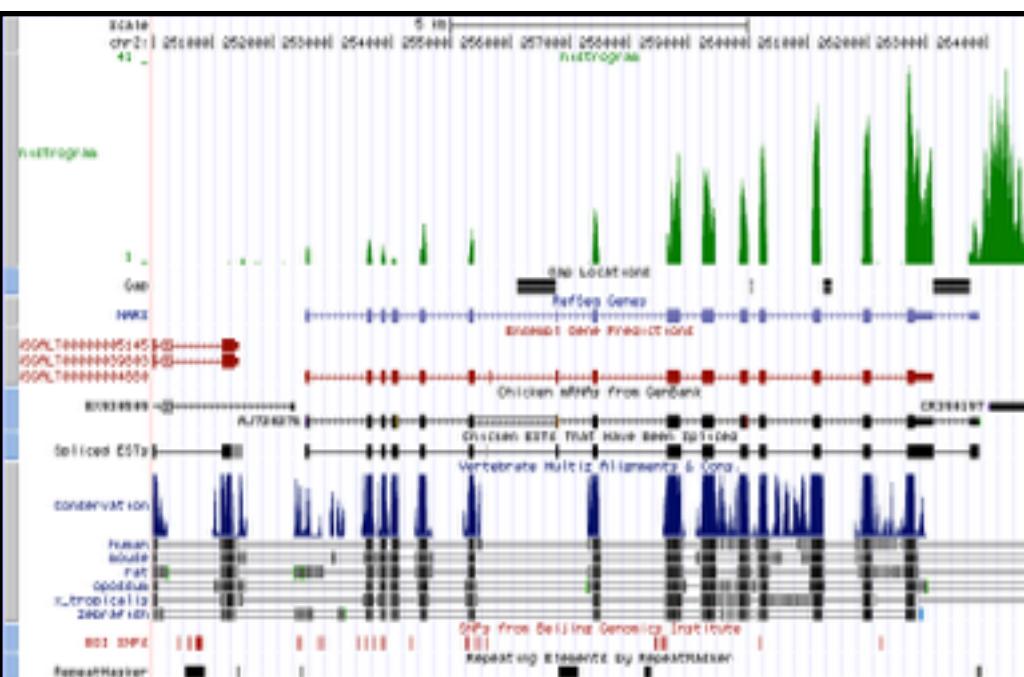


motivation

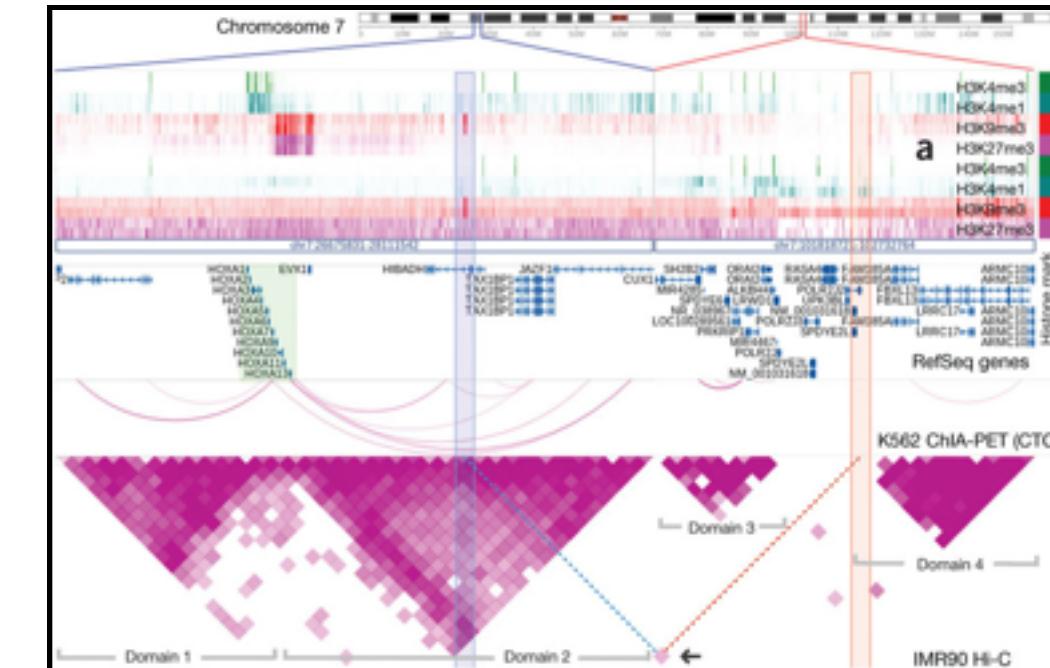
<http://bit.ly/2gufWsB>

```
git clone https://github.com/hyounesy/cass2017 vis.git
```

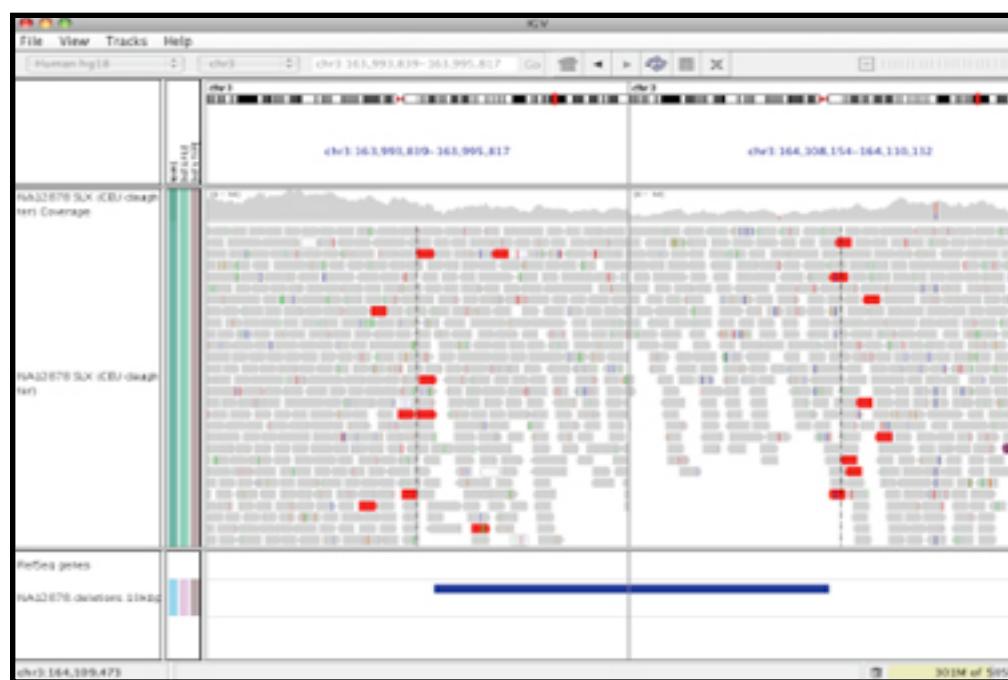
advanced



UCSC Genome Browser



WashU Epigenome Browser



Integrated Genome Viewer

basic

complex

ease of use

easy

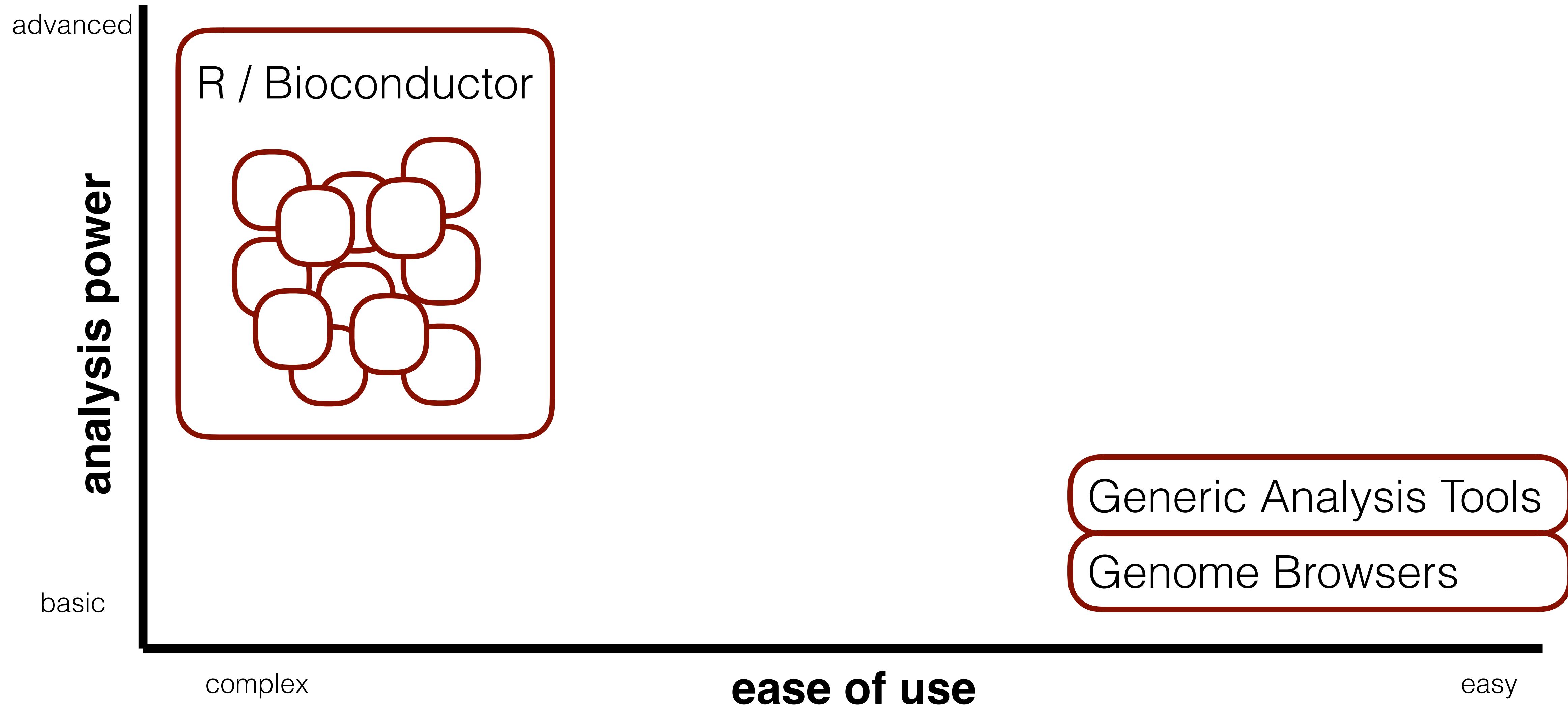
Generic Analysis Tools

Genome Browsers

motivation

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git



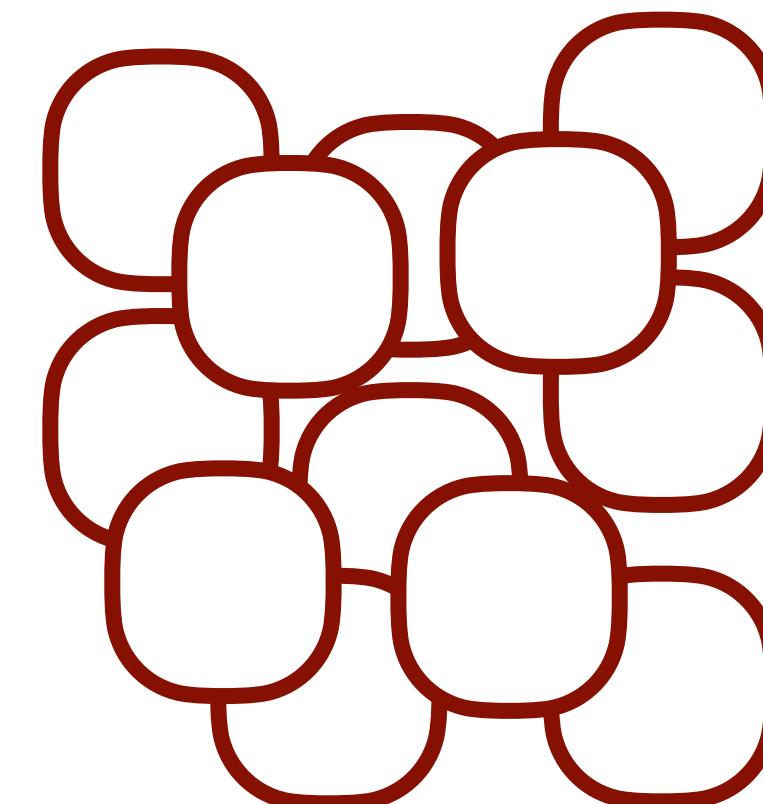
motivation

advanced

analysis power

basic

R / Bioconductor



complex

ease of use

easy

Packages found under Sequencing:

Package	Maintainer	Title
AIMS	Eric R. Paquet	AIMS : Absolute Assignment of Breast Cancer Intrinsic Molecular Subtype
ALDEX2	Greg Gloor	Analysis of differential abundance taking sample variation into account
AllelicImbalance	Jesper R. Gadin	Investigates allele specific expression
ampliQueso	Michał Okoniewski	Analysis of amplicon enrichment panels
AnnotationDbi	Bioconductor Package Maintainer	Annotation Database Interface
anota	Ola Larsson	ANalysis Of Translational Activity (ANOTA).
ArrayExpressHTS	Angela Goncalves, Andrew Tikhonov	ArrayExpress High Throughput Sequencing Processing Pipeline
BADER	Andreas Neudecker	Bayesian Analysis of Differential Expression in RNA Sequencing Data
ballgown	Alyssa Frazee	Flexible, isoform-level differential expression analysis
bamsignals	Alessandro Mammana	Extract read count signals from bam files

Bioconductor Repository

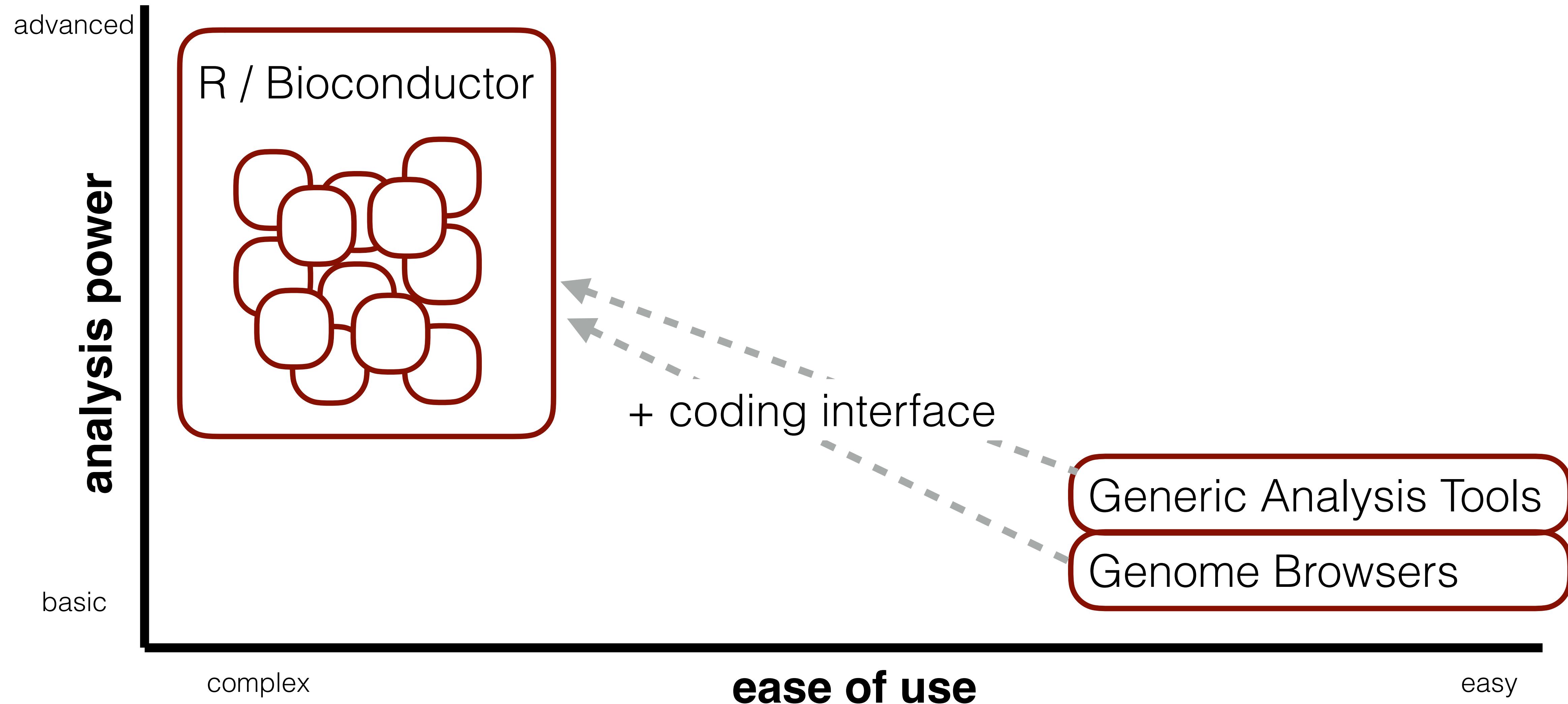
Generic Analysis Tools

Genome Browsers

motivation

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git



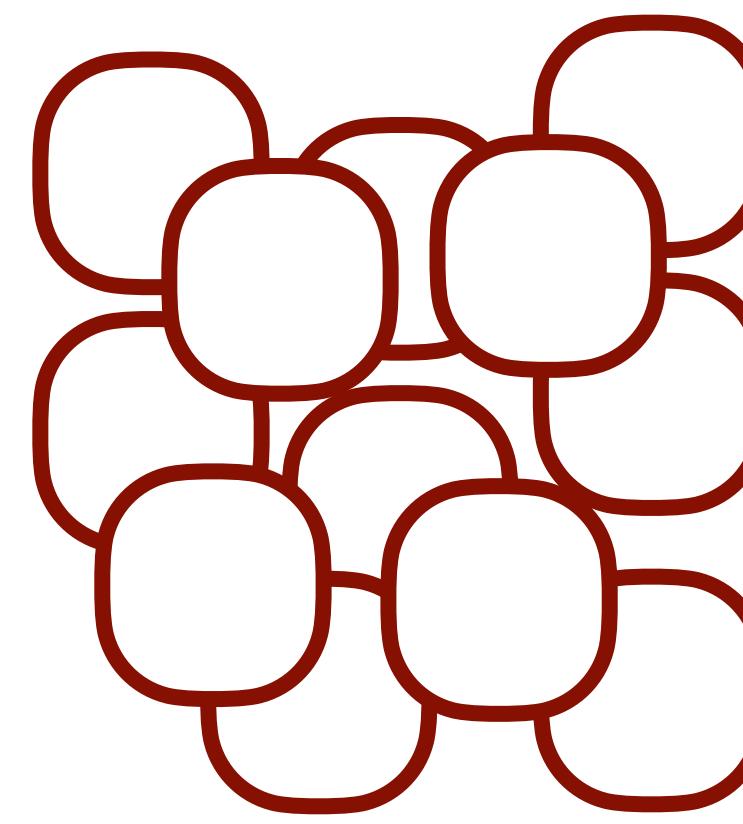
motivation

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git

advanced

R / Bioconductor



analysis power

basic

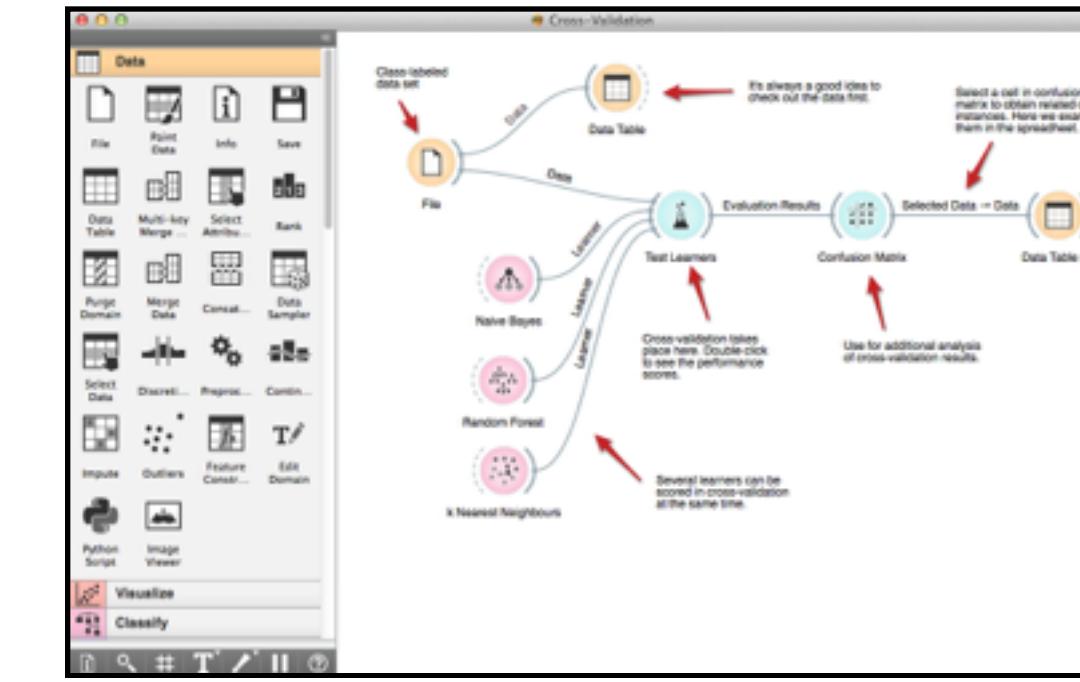
complex

ease of use

easy



KNIME



Orange

+ coding interface

Generic Analysis Tools
Genome Browsers

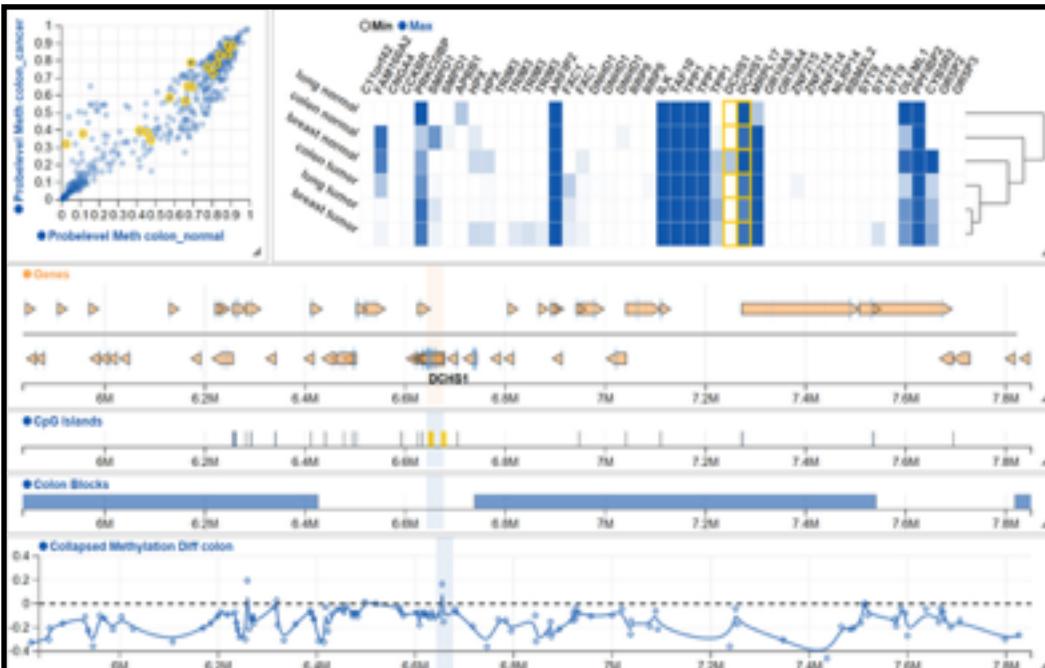
motivation

<http://bit.ly/2gufWsB>

```
git clone https://github.com/hyounesy/cass2017 vis.git
```

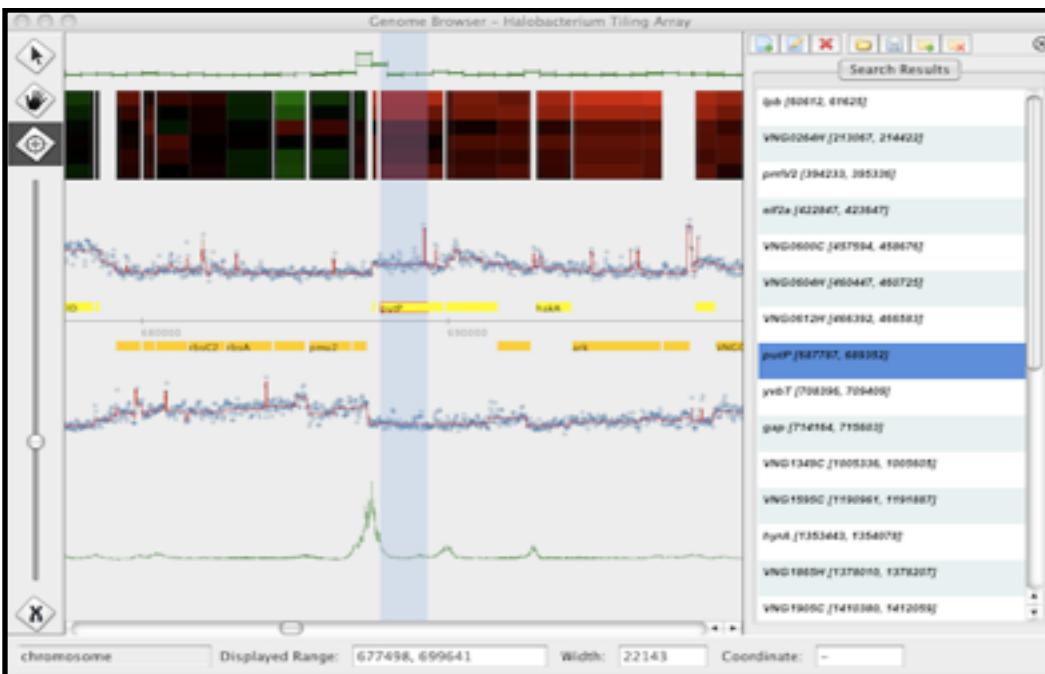
advanced

R / Bioconductor



Epiviz

basic



Gaggle Genome Browser

KNIME

Orange

+ coding interface

Generic Analysis Tools

Genome Browsers

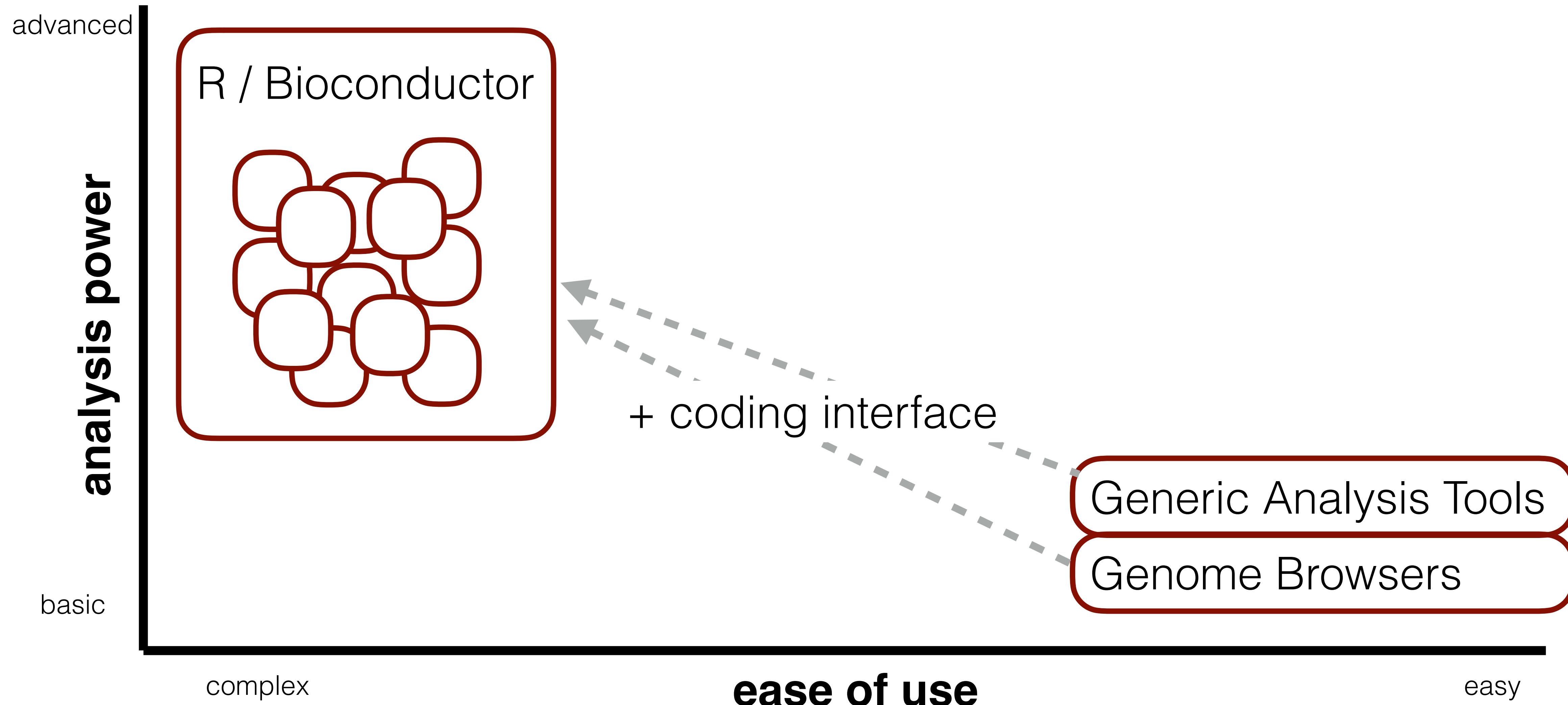
ease of use

easy

motivation

<http://bit.ly/2gufWsB>

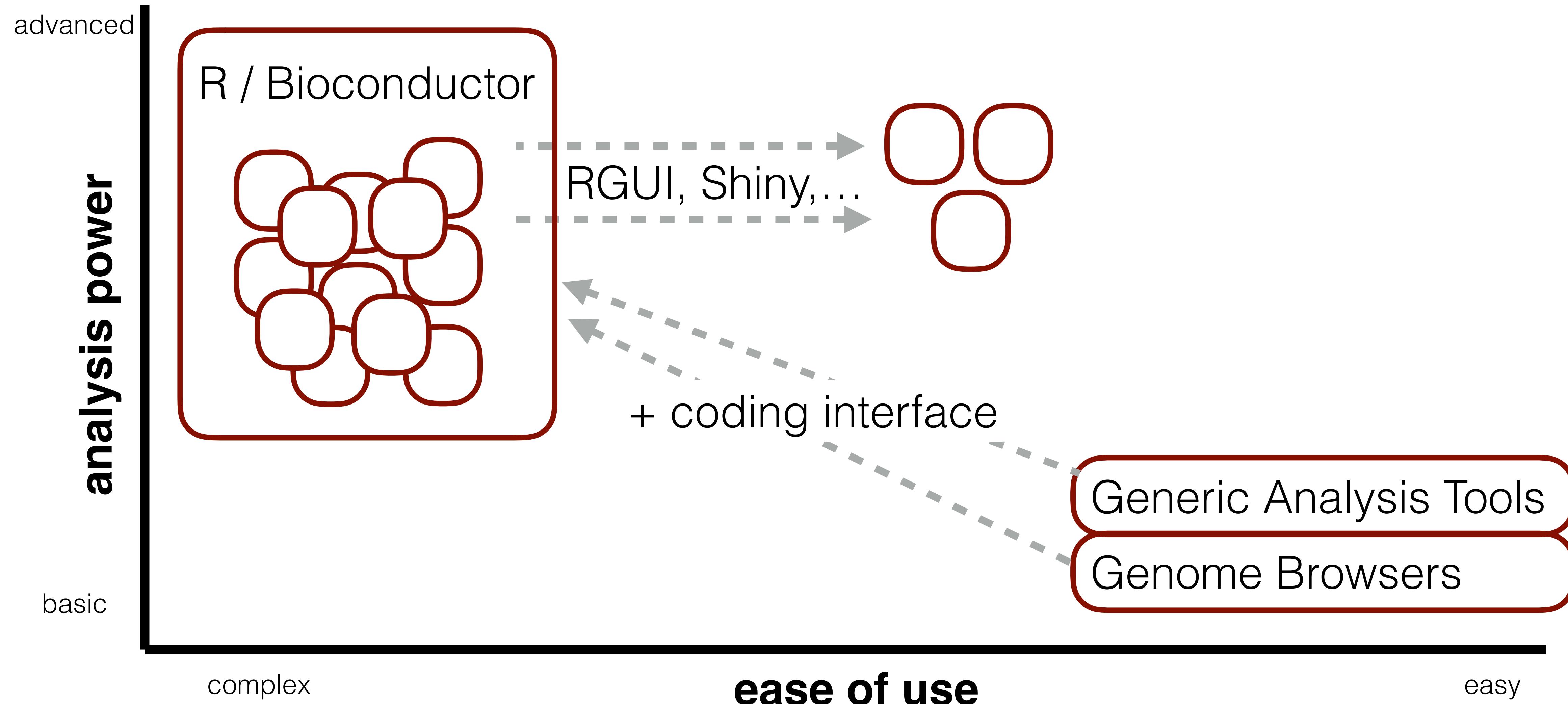
git clone https://github.com/hyounesy/cass2017_vis.git



motivation

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git



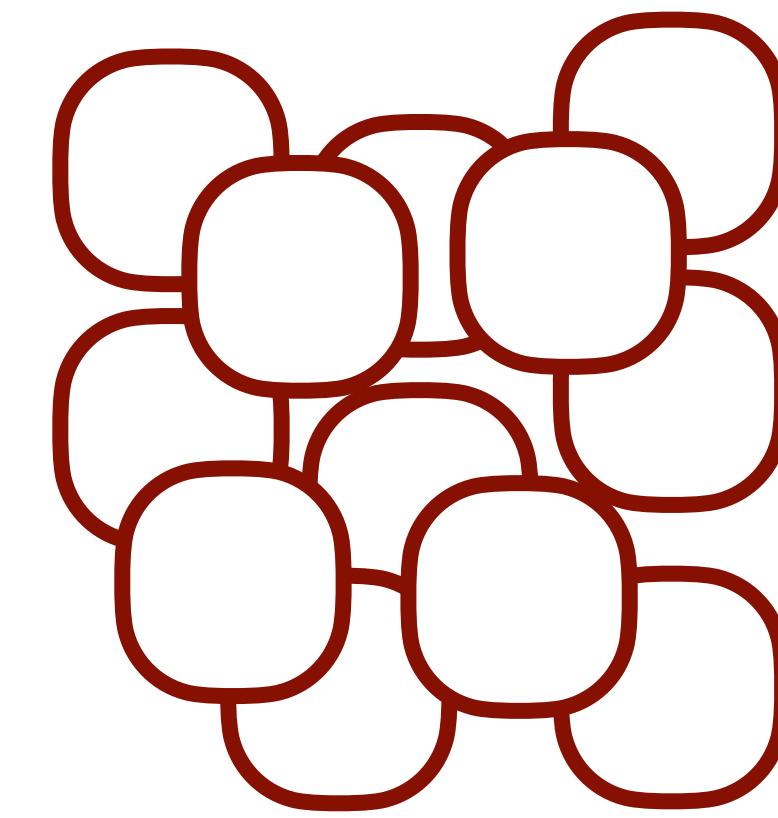
motivation

advanced

analysis power

basic

R / Bioconductor



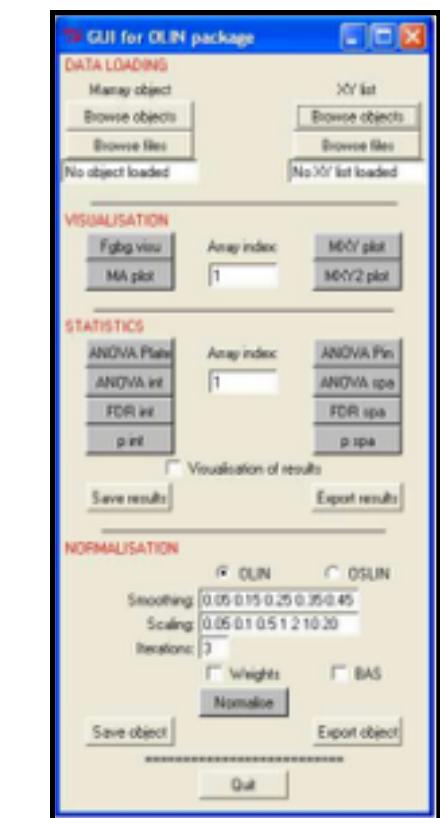
RGUI, Shiny, ...

+ coding interface

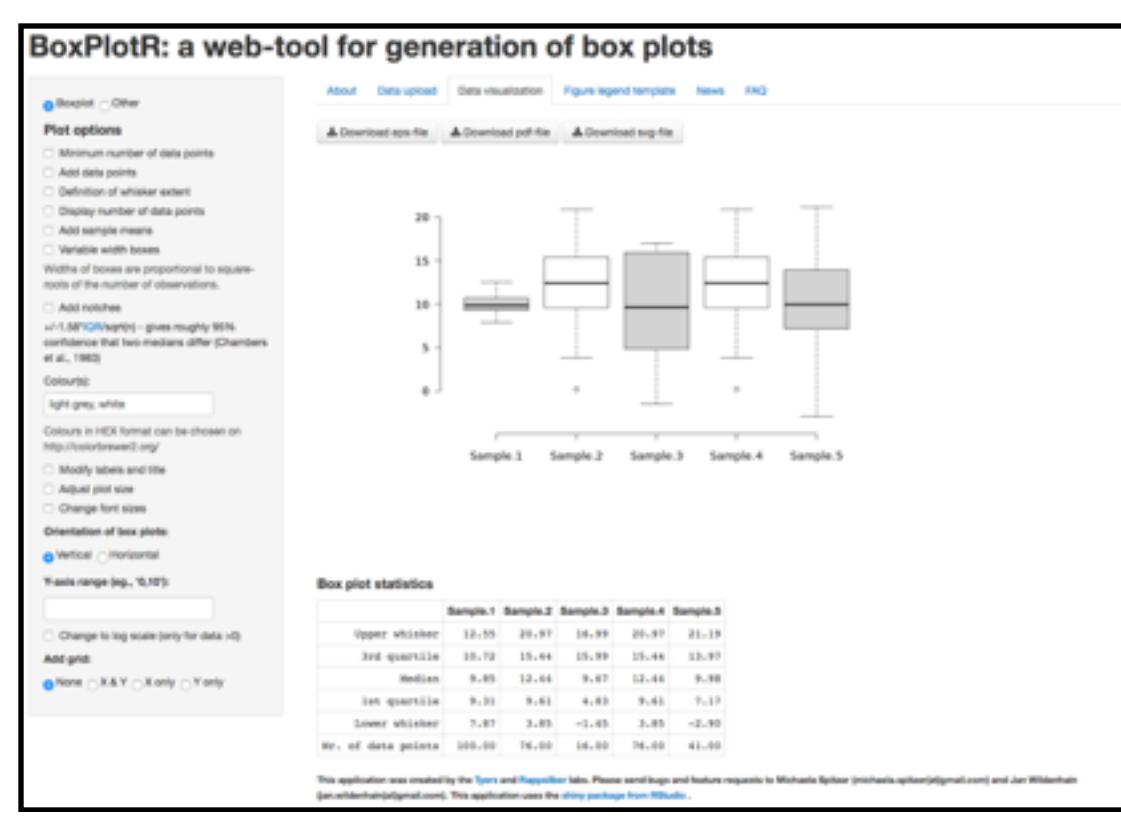
complex

ease of use

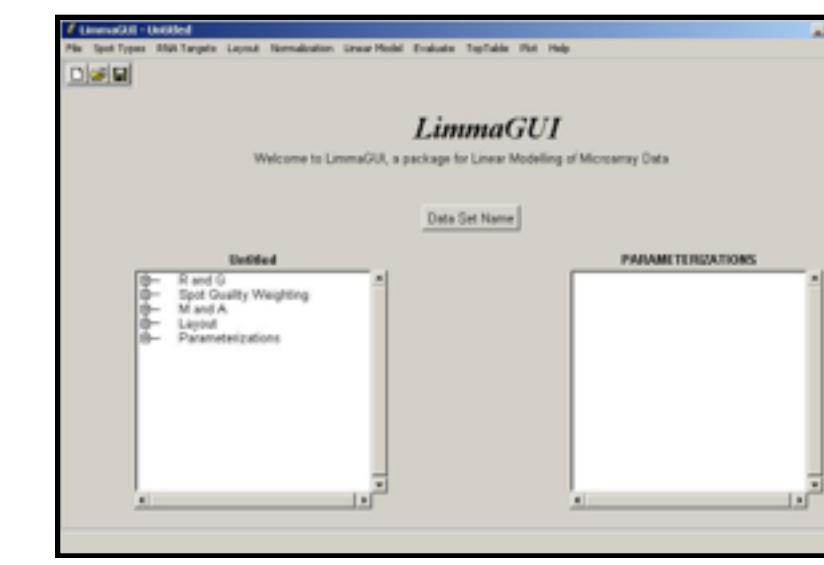
easy



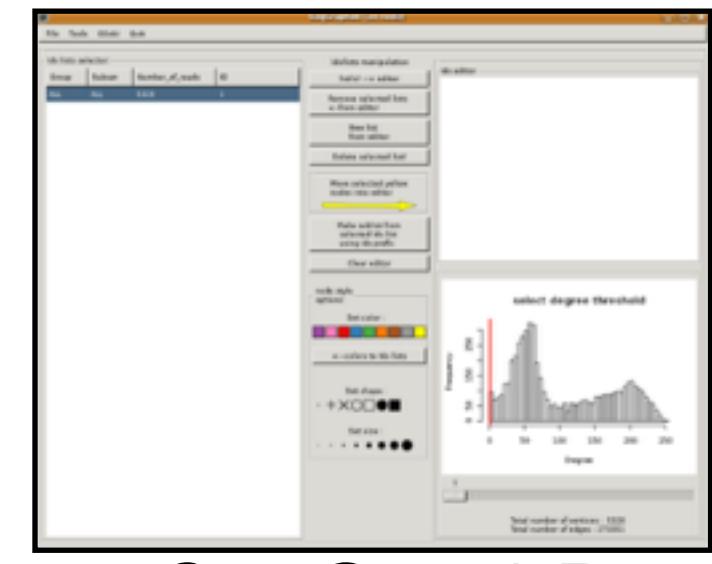
OLINgui



BoxPlotR



LimmaGUI



SeqGraphR

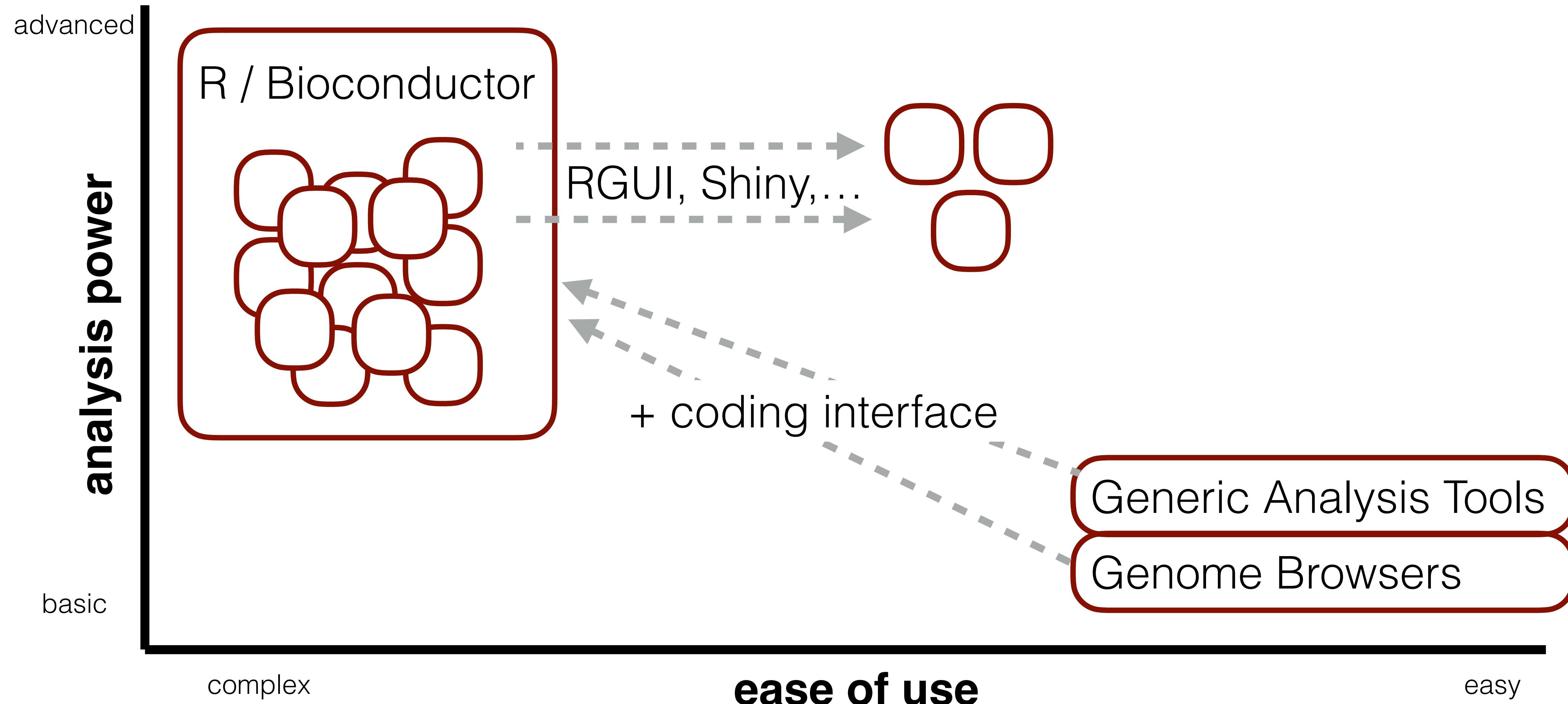
Generic Analysis Tools

Genome Browsers

motivation

<http://bit.ly/2gufWsB>

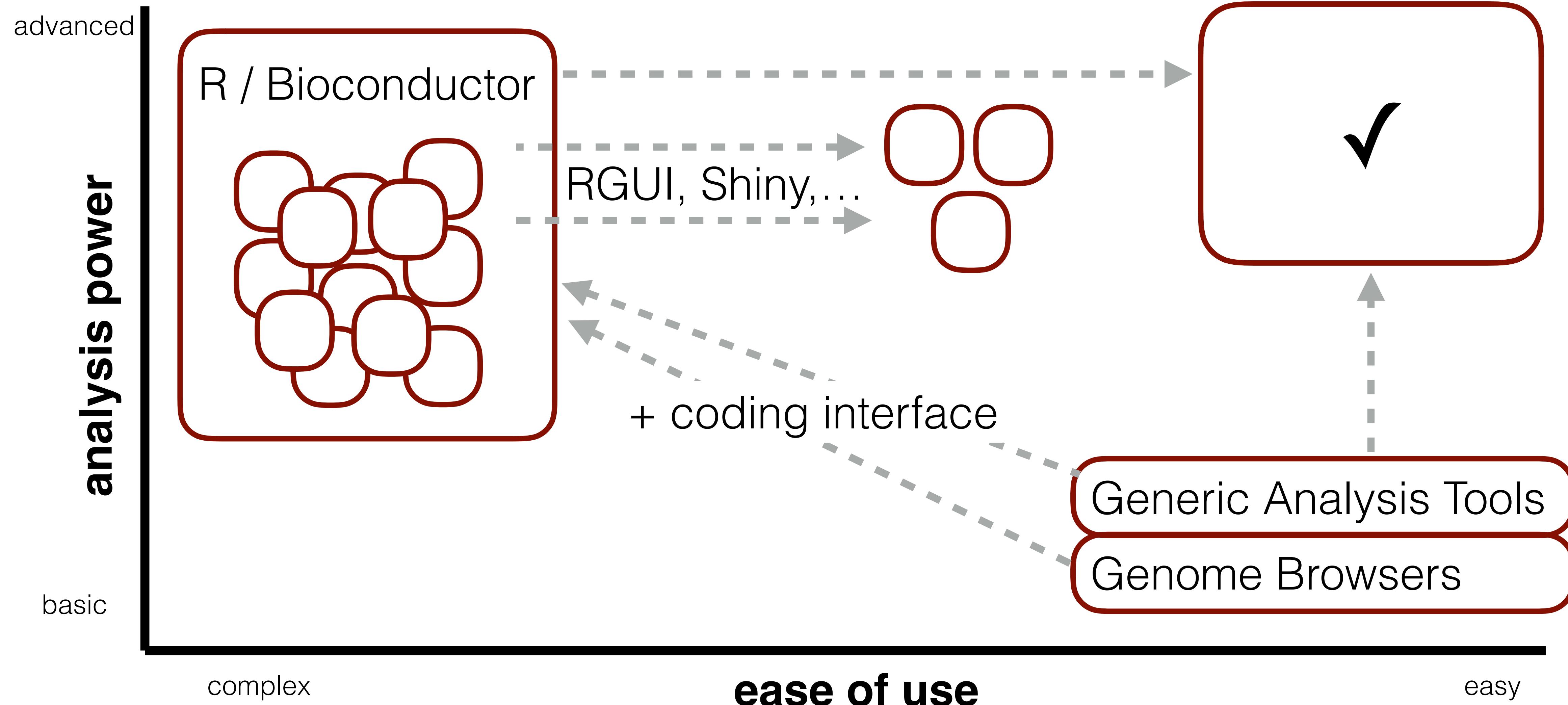
git clone https://github.com/hyounesy/cass2017_vis.git



motivation

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git



<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git

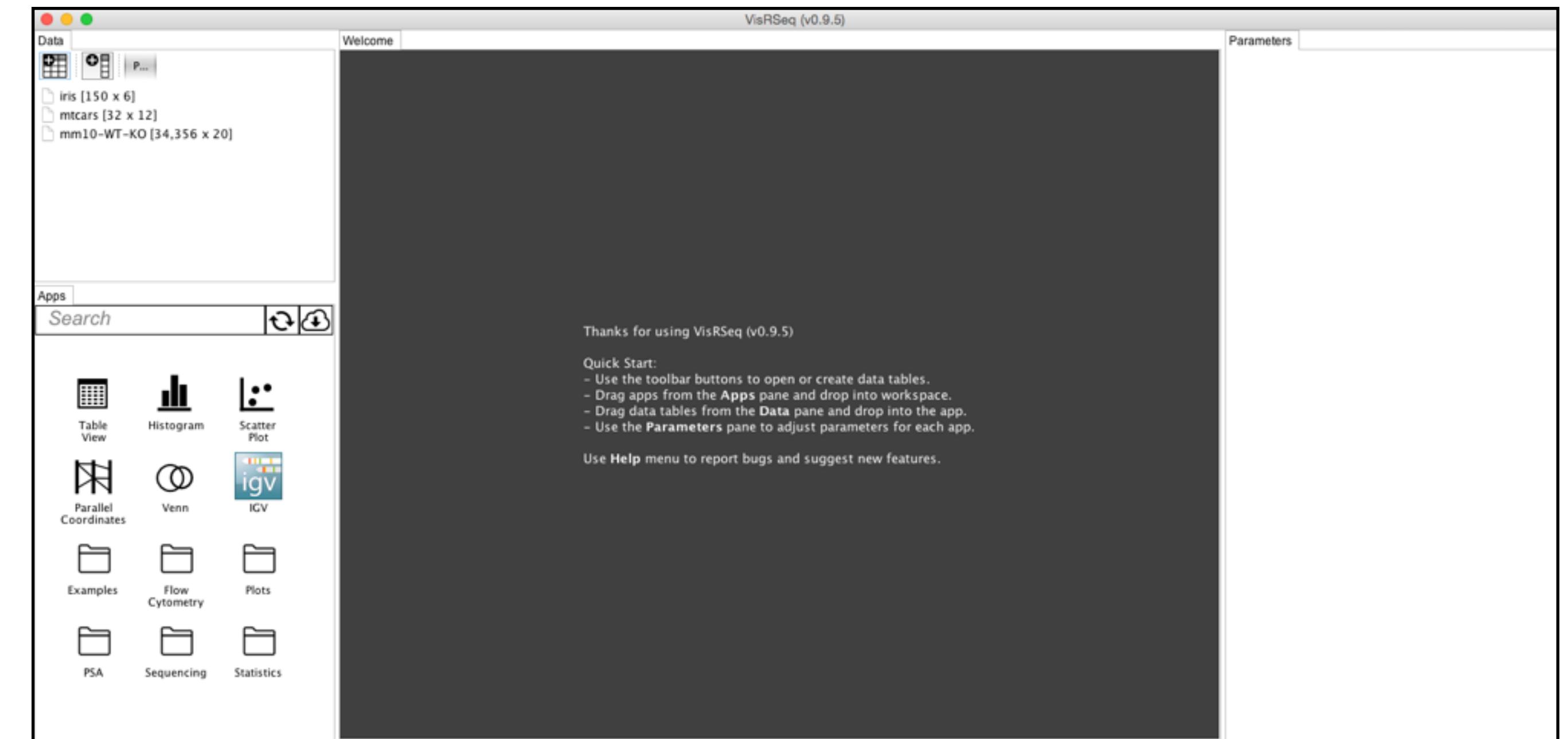
VisRseq

interface

- Installation Requirements:
 - Java SE 7
 - R
- Auto installs required packages
- Open Beta: <http://visrseq.github.io>

<http://bit.ly/2gufWsB>

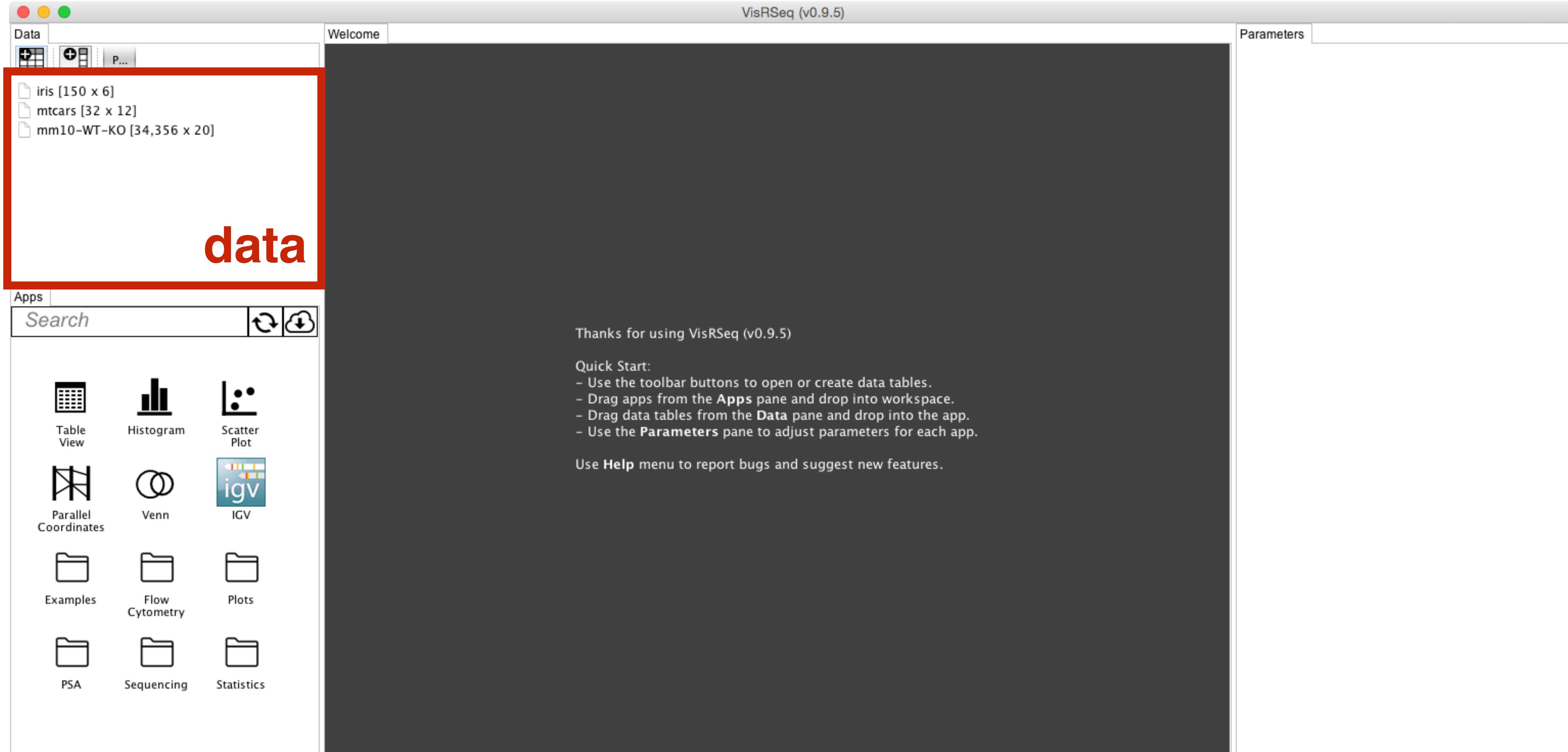
git clone https://github.com/hyounesy/cass2017_vis.git



interface: data

<http://bit.ly/2gufWsB>

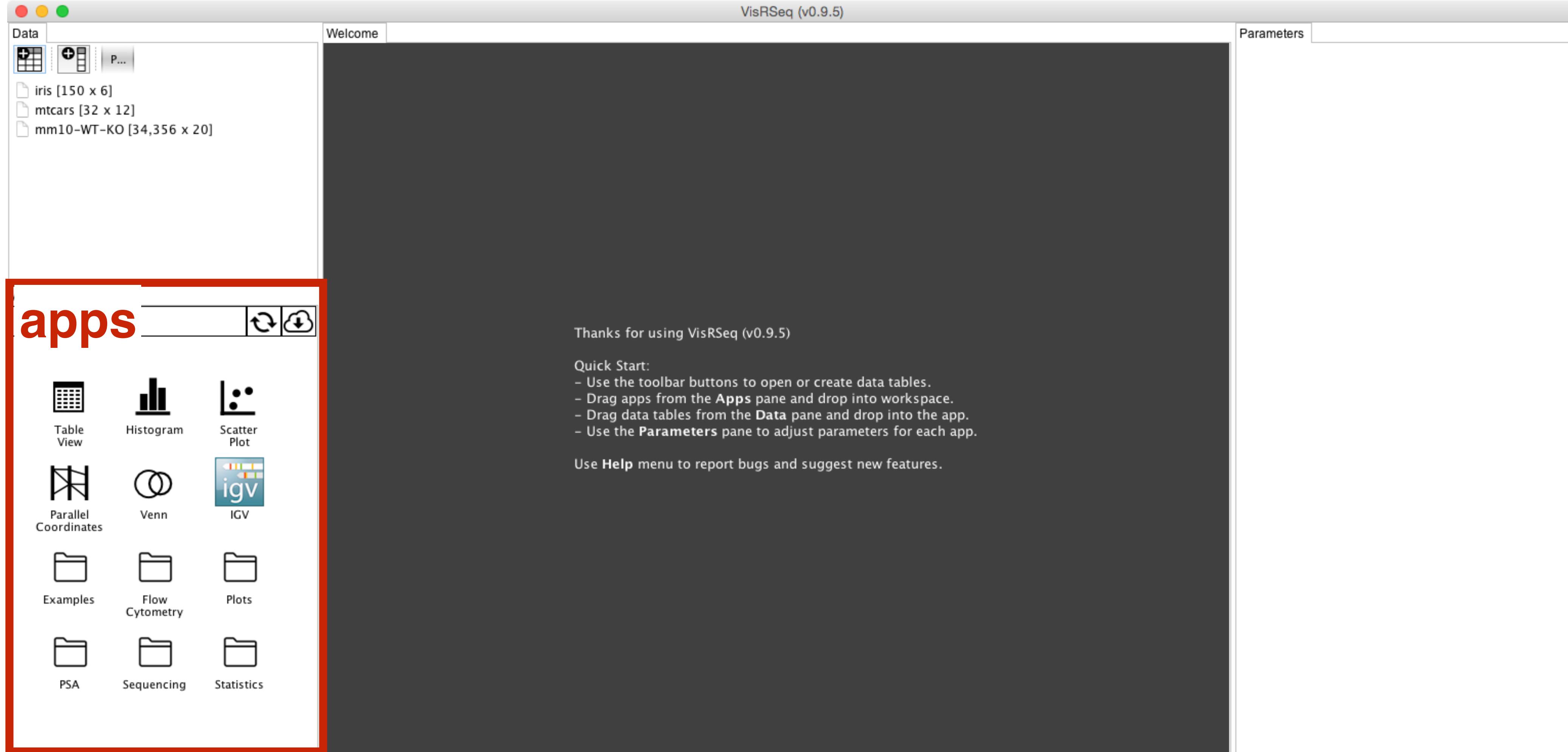
git clone https://github.com/hyounesy/cass2017_vis.git



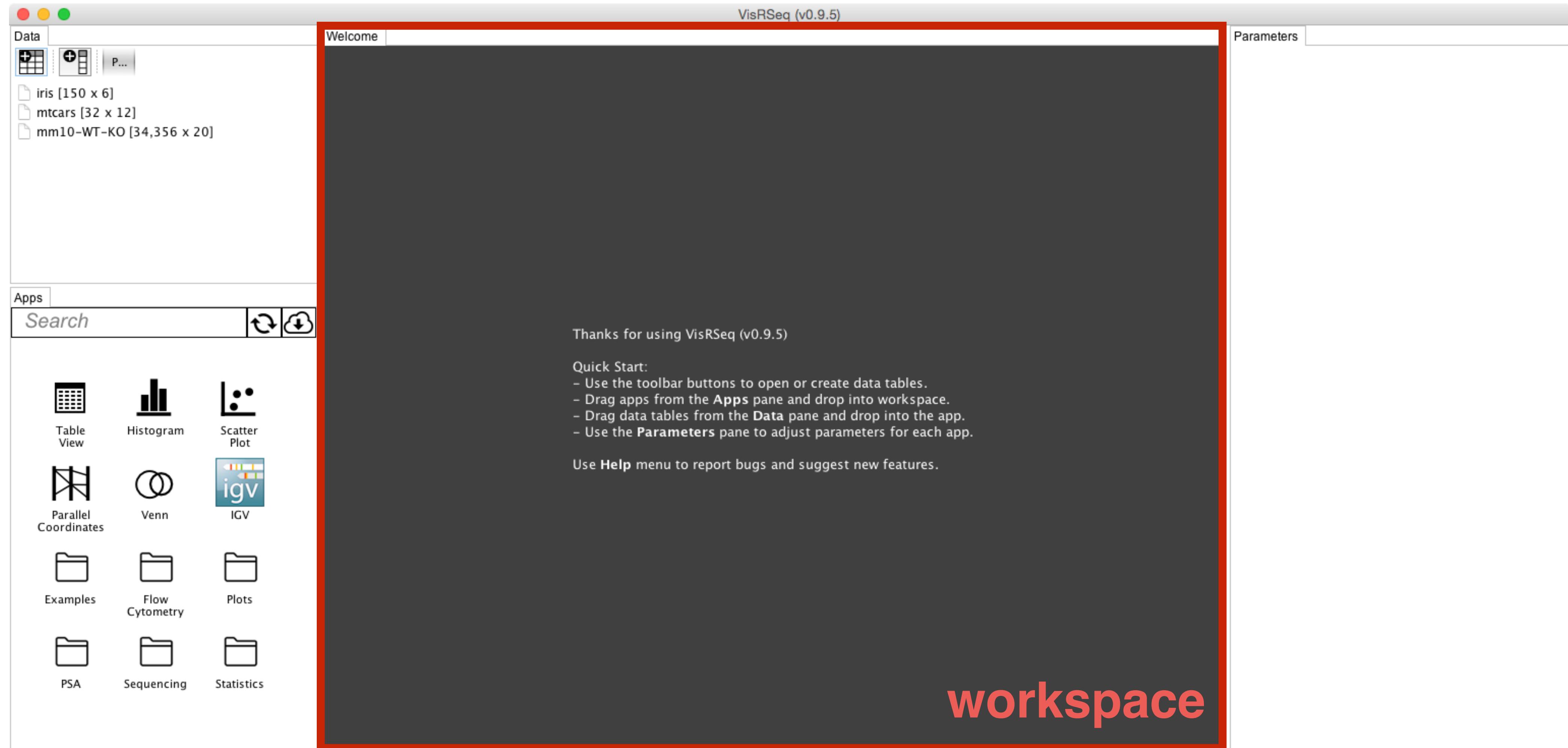
interface: apps

<http://bit.ly/2gufWsB>

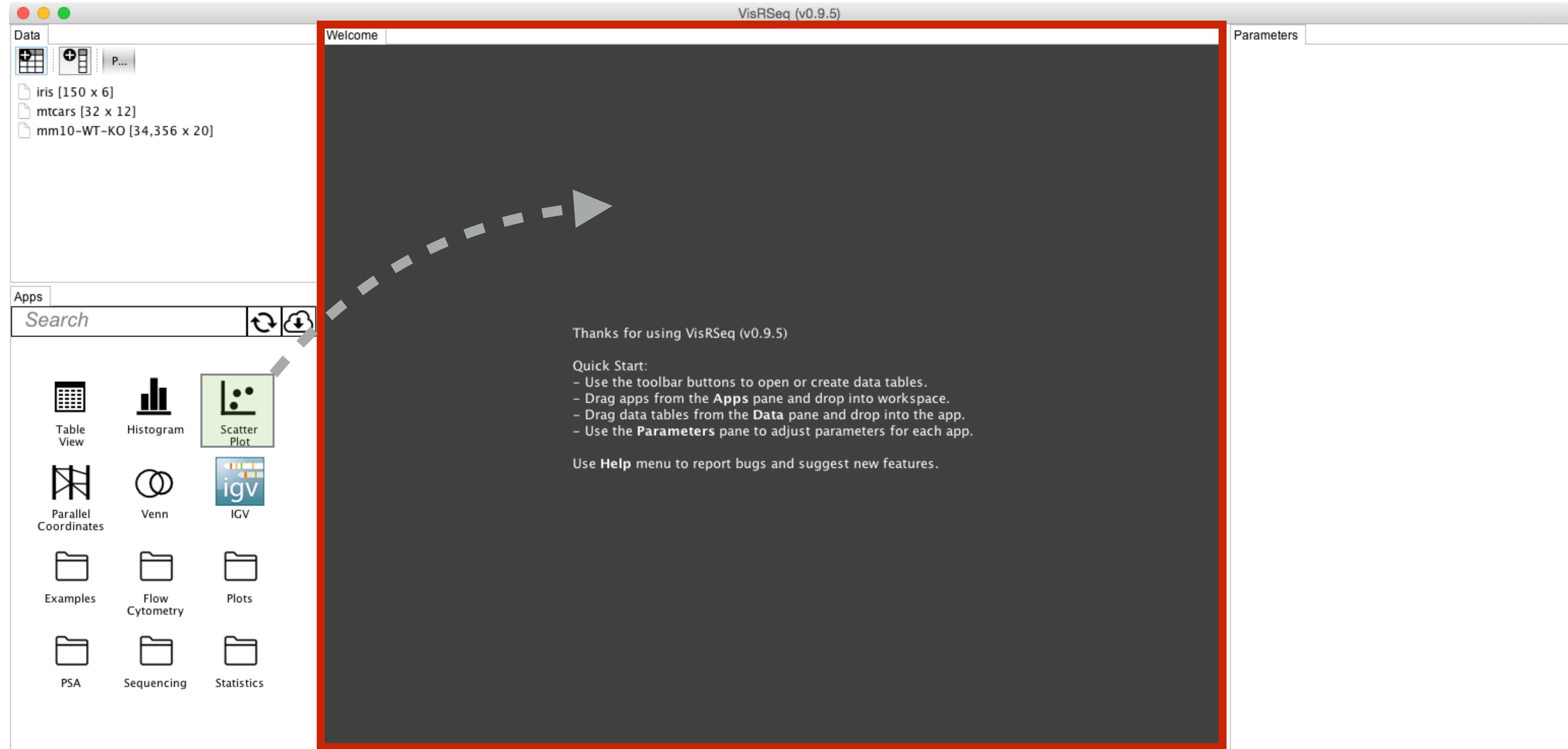
git clone https://github.com/hyounesy/cass2017_vis.git



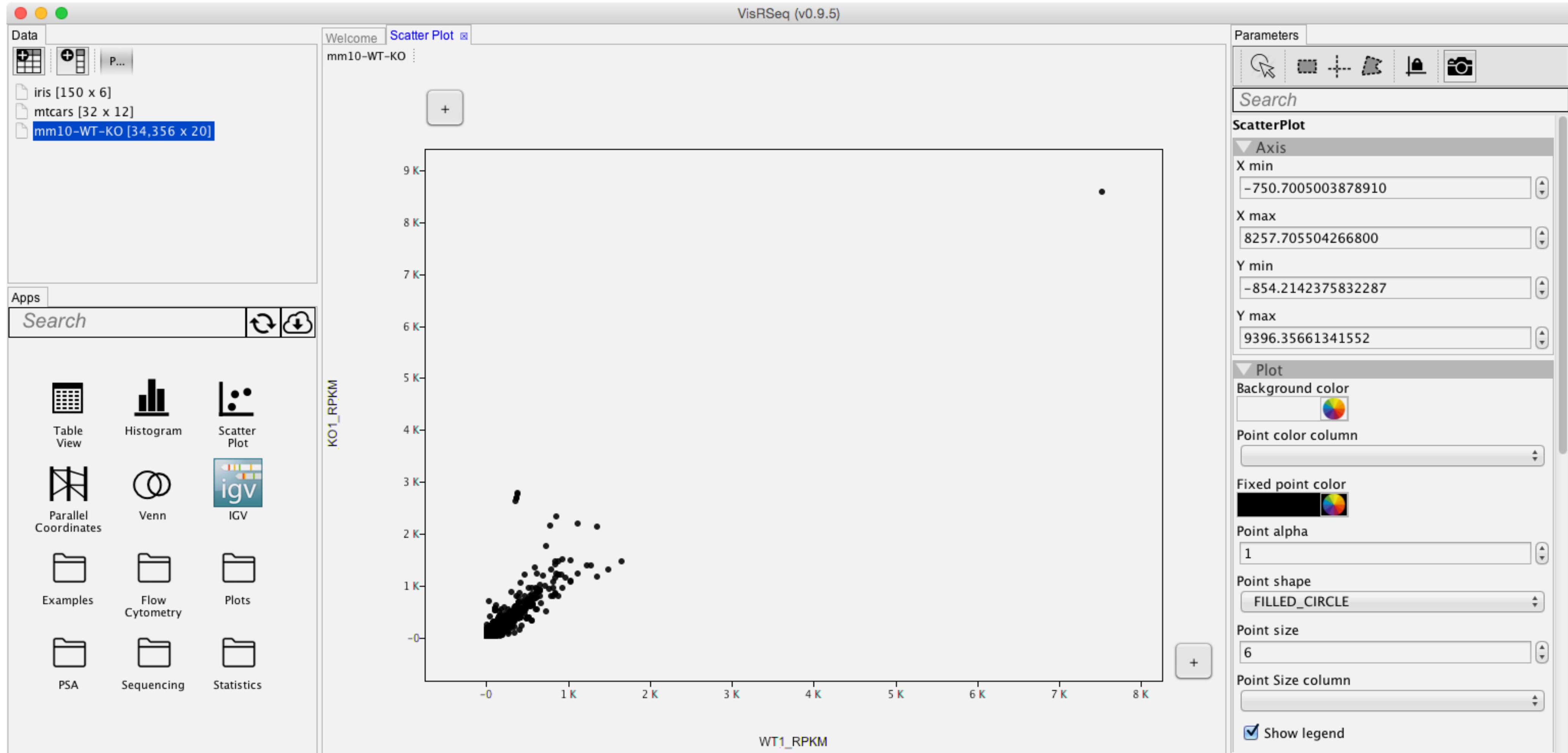
interface: workspace



interface: workspace

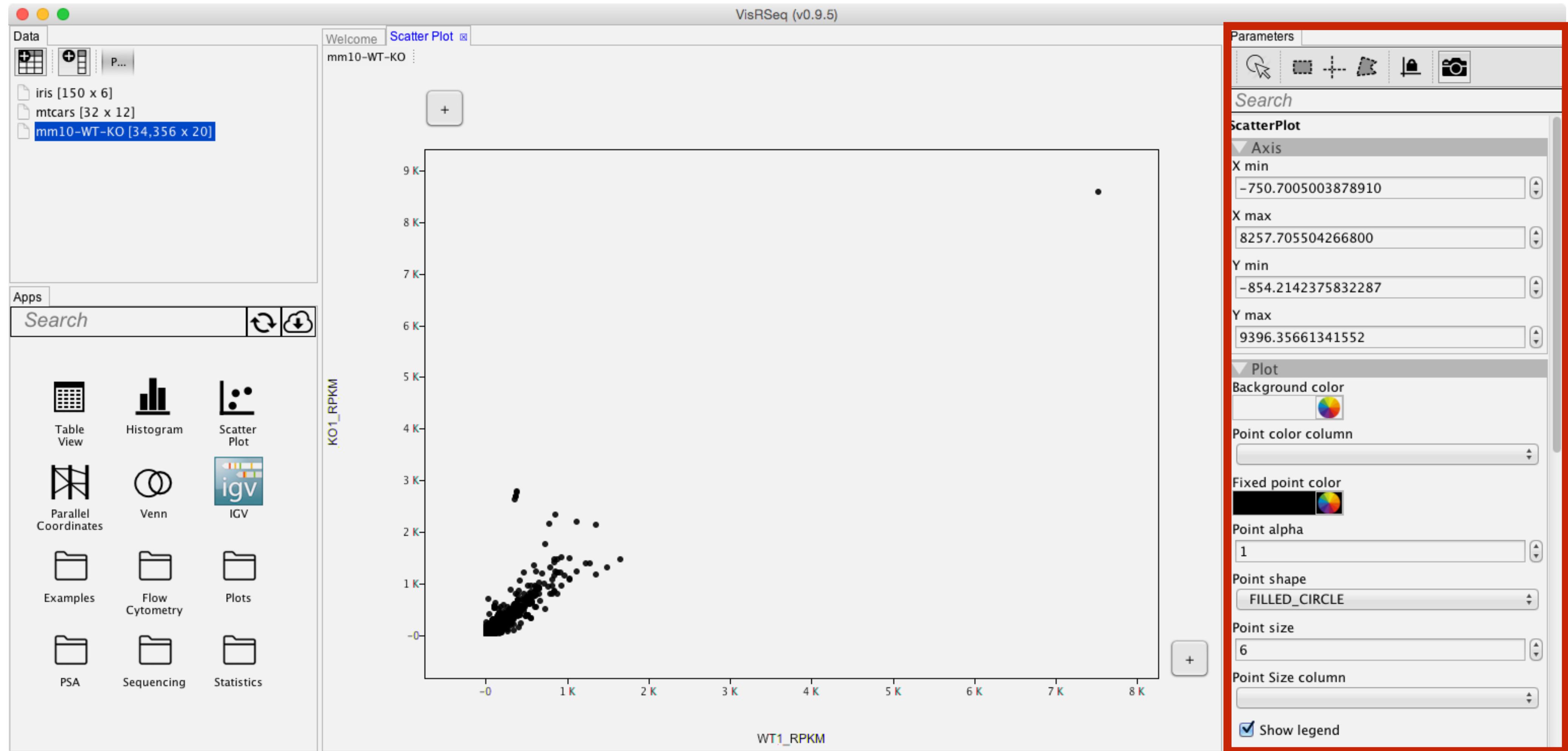


interface: workspace



interface: parameters

parameters



<http://bit.ly/2gufWsB>

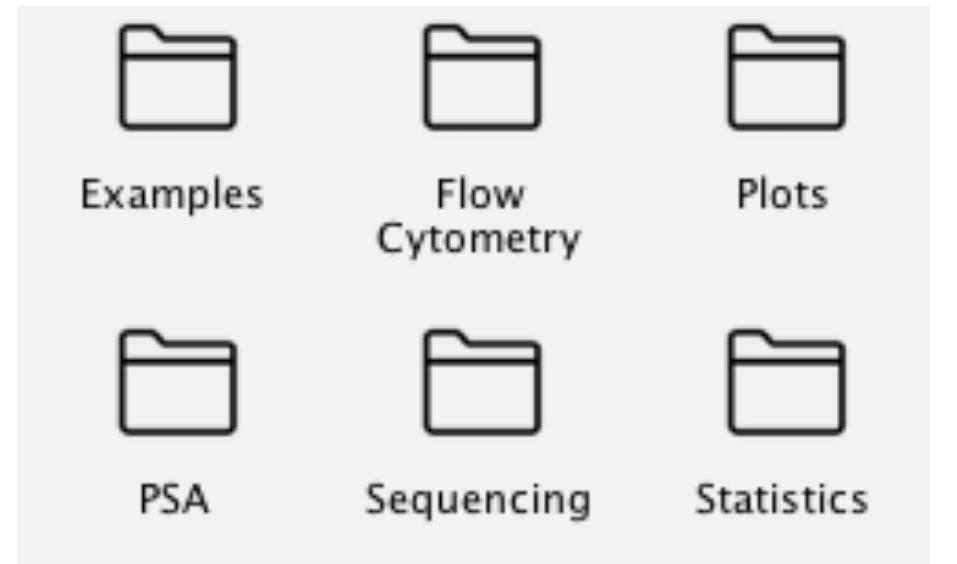
git clone https://github.com/hyounesy/cass2017_vis.git

R - apps

R Apps

<http://bit.ly/2gufWsB>

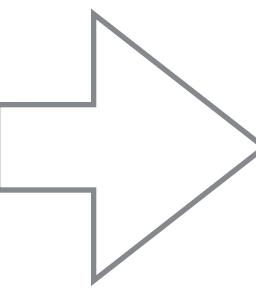
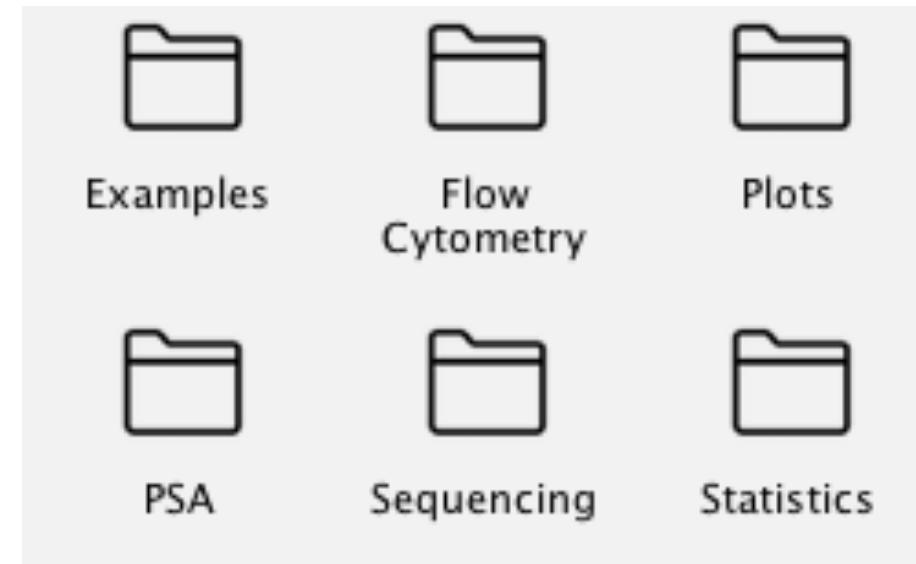
git clone https://github.com/hyounesy/cass2017_vis.git



R Apps

<http://bit.ly/2gufWsB>

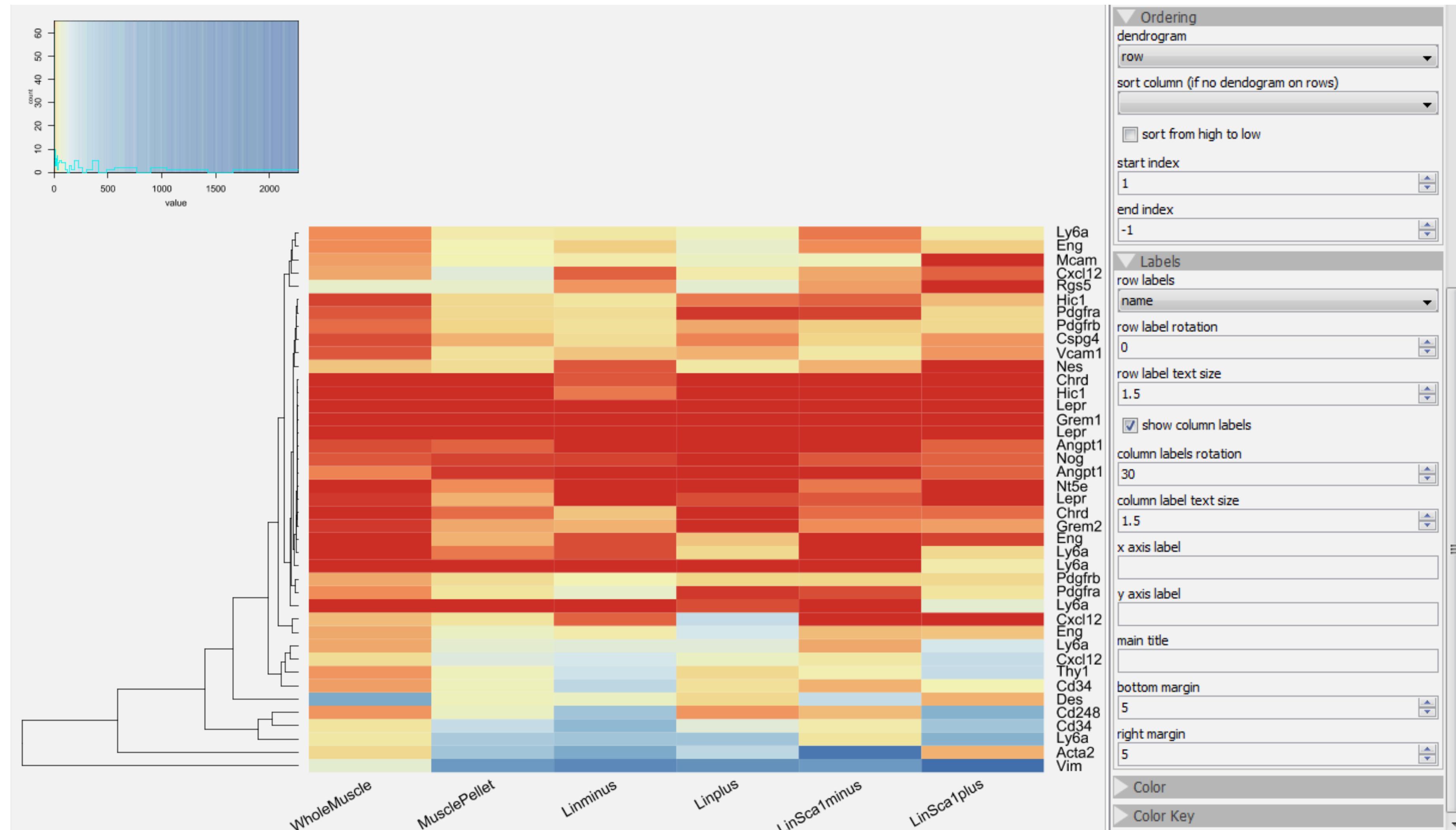
git clone https://github.com/hyounesy/cass2017_vis.git



The main interface displays several categories of R apps:

- [Plots]**: BarPlot, Boxplot, Correlogram, DensityPlot, DensityPlot2, DensityPlot3, heatmap, MosaicPlot, OmicCircos, PieChart, Scatterplot 3D, Scatterplot ggplot, ScreePlot, ScreePlot2, Stacked Histogram.
- [Statistics]**: Clustering, Dimensional Analysis, Distributions, Models, Summary, Tests.
- [RNA-seq]**: baySeq, Compute RLE, DESeq, DESeq2, edgeR, NOISeq, sSeq, timeClip, tweeDEseq, VolcanoPlot.
- [Flow Cytometry]**: flowClust, flowMeans, SamSPECTRAL, summary.

R – Apps example: heatmap



Anatomy of an R app

R code

kmeans.R

```
source("visrutils.R")

visr.applyParameters()

cluster_data<-subset(visr.input, select = param.columns)

output.clusterid <- kmeans(cluster_data, param.k,
                           algorithm = param.algorithm)$cluster
plot(cluster_data, main = param.plot.title,
      col = as.integer(output.clusterid))
```

parameters

kmeans.json

```
{ "label": "kmeans", "categories": [
  { "variables": {
      "param.columns": {"type": "multi-column-numerical" },
      "param.k": {"type": "int", "default": 3 },
      "param.algorithm": {"items": ["Hartigan-Wong", "Lloyd",
                                    "Forgy", "MacQueen"]},
      "param.plot.title": {"default": "kmeans result" },
      "output.clusterid": {"type": "output-column" }
    }
  }
]}
```

Anatomy of an R app

R code

kmeans.R

```
source("visrutils.R")

visr.applyParameters()

cluster_data<-subset(visr.input, select = param.columns)

output.clusterid <- kmeans(cluster_data, param.k,
                            algorithm = param.algorithm)$cluster
plot(cluster_data, main = param.plot.title,
      col = as.integer(output.clusterid))
```

parameters

kmeans.json

```
{ "label": "kmeans", "categories": [
  { "variables": {
    "param.columns": {"type": "multi-column-numerical"}, 
    "param.K": {"type": "int", "default": 3}, 
    "param.algorithm": {"items": ["Hartigan-Wong", "Lloyd", "Forgy", "MacQueen"]}, 
    "param.plot.title": {"default": "kmeans result"}, 
    "output.clusterid": {"type": "output-column"}
  }}
]}
```

Anatomy of an R app

R code

kmeans.R

```
source("visrutils.R")

visr.applyParameters()

cluster_data<-subset(visr.input, select = param.columns)

output.clusterid <- kmeans(cluster_data, param.k
                            algorithm = param.algorithm$cluster
plot(cluster_data, main = param.plot.title,
      col = as.integer(output.clusterid))
```

parameters

kmeans.json

```
{ "label": "kmeans", "categories": [
  { "variables": {
    "param.columns": {"type": "multi-column-numerical" },
    "param.k": {"type": "int", "default": 3 },
    "param.algorithm": {"items": ["Hartigan-Wong", "Lloyd",
                                 "Forgy", "MacQueen"]},
    "param.plot.title": {"default": "kmeans result" },
    "output.clusterid": {"type": "output-column" }
  }}
]}
```

Anatomy of an R app

R code

kmeans.R

```
source("visrutils.R")

visr.applyParameters()

cluster_data<-subset(visr.input, select = param.columns)

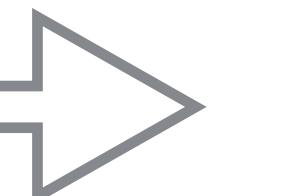
output.clusterid <- kmeans(cluster_data, param.k,
                           algorithm = param.algorithm)$cluster

plot(cluster_data, main = param.plot.title,
     col = as.integer(output.clusterid))
```

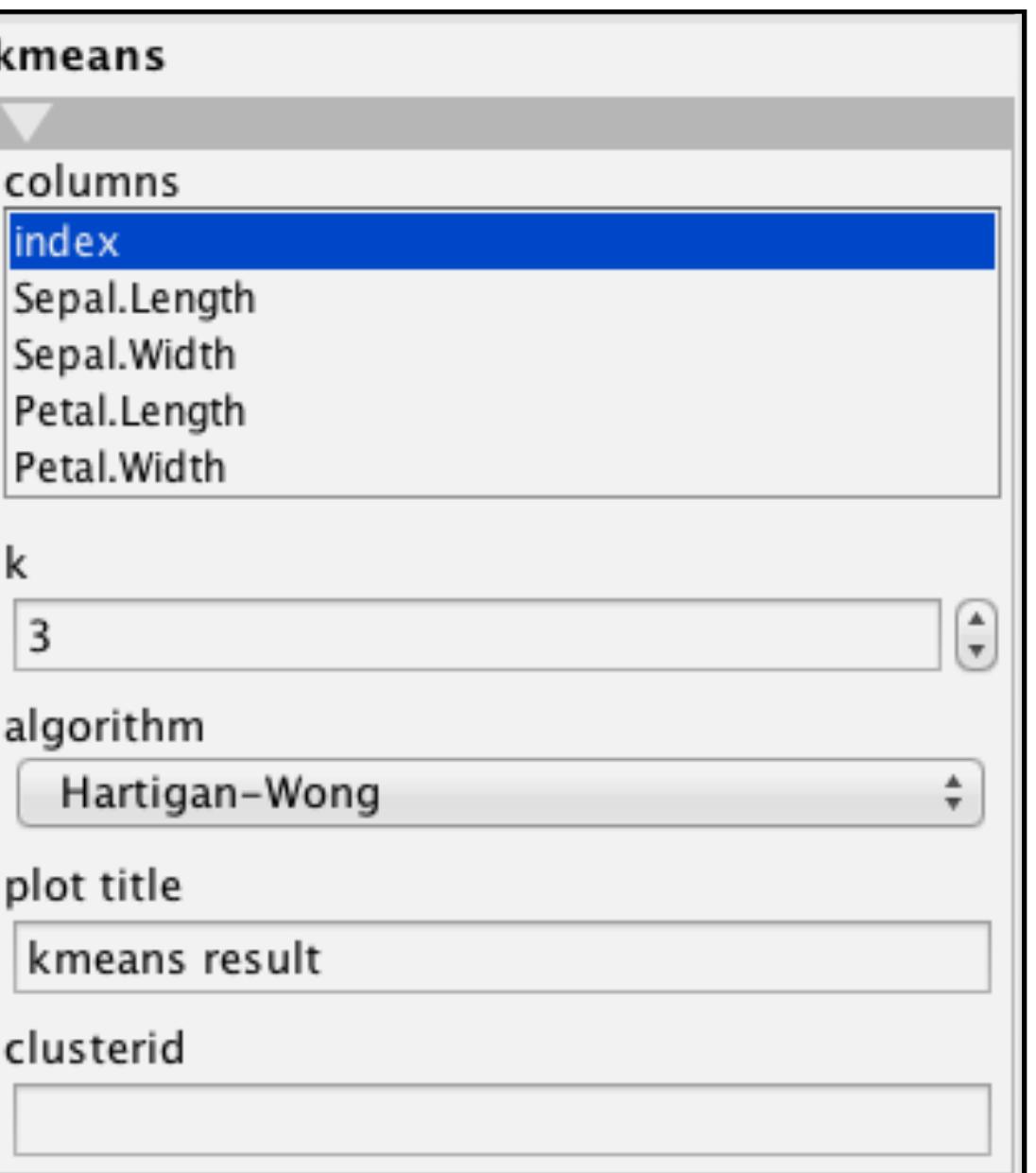
parameters

kmeans.json

```
{ "label": "kmeans", "categories": [
  { "variables": {
      "param.columns": {"type": "multi-column-numerical" },
      "param.k": {"type": "int", "default": 3 },
      "param.algorithm": {"items": ["Hartigan-Wong", "Lloyd",
                                    "Forgy", "MacQueen"]},
      "param.plot.title": {"default": "kmeans result"},
      "output.clusterid": {"type": "output-column"}
    }
]}]
```



auto generated UI



Anatomy of an R app

R code

kmeans.R

```
source("visrutils.R")

visr.applyParameters()

cluster_data<-subset(visr.input, select = param.columns)

output.clusterid <- kmeans(cluster_data, param.k,
                           algorithm = param.algorithm)$cluster

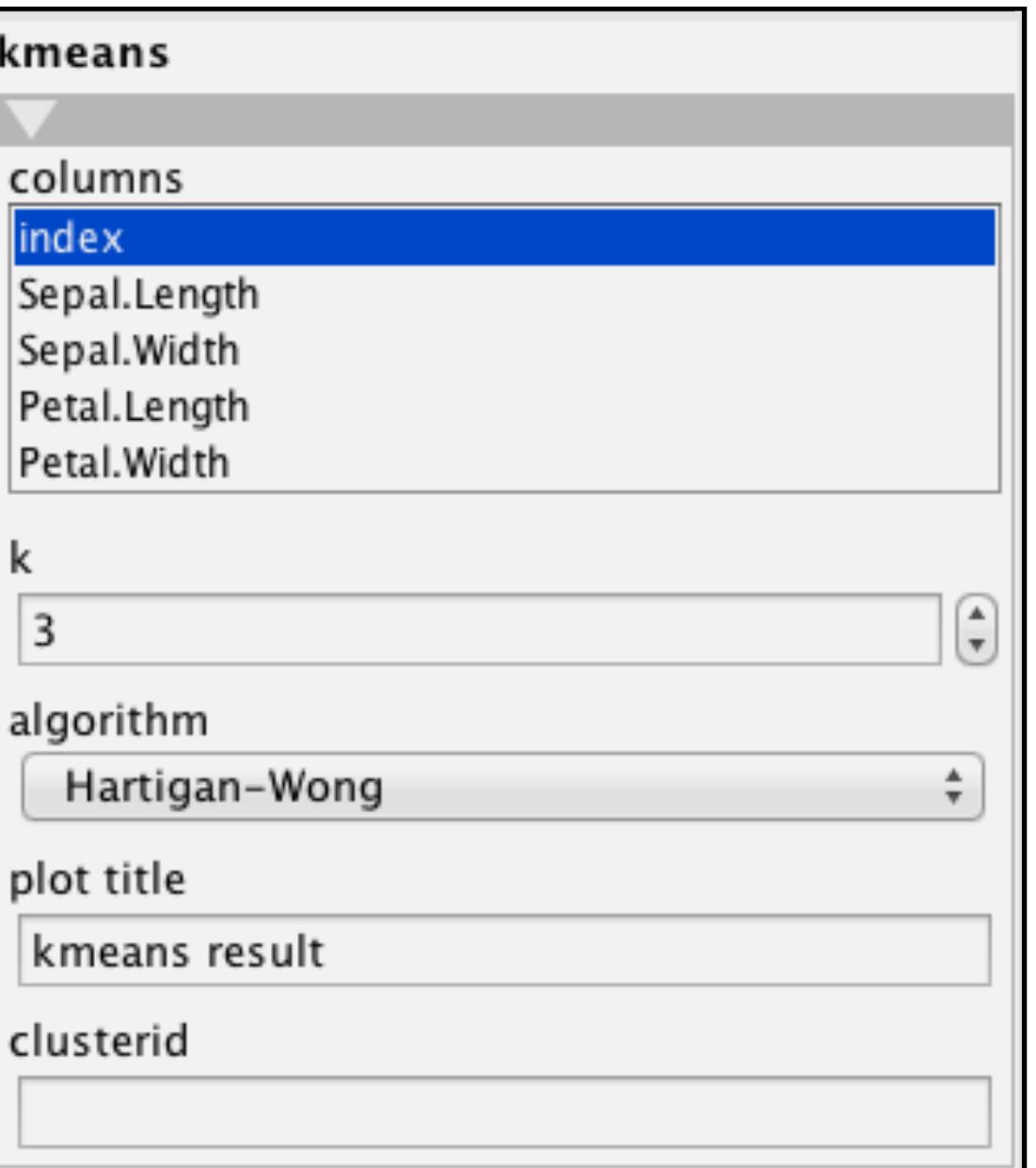
plot(cluster_data, main = param.plot.title,
     col = as.integer(output.clusterid))
```

parameters

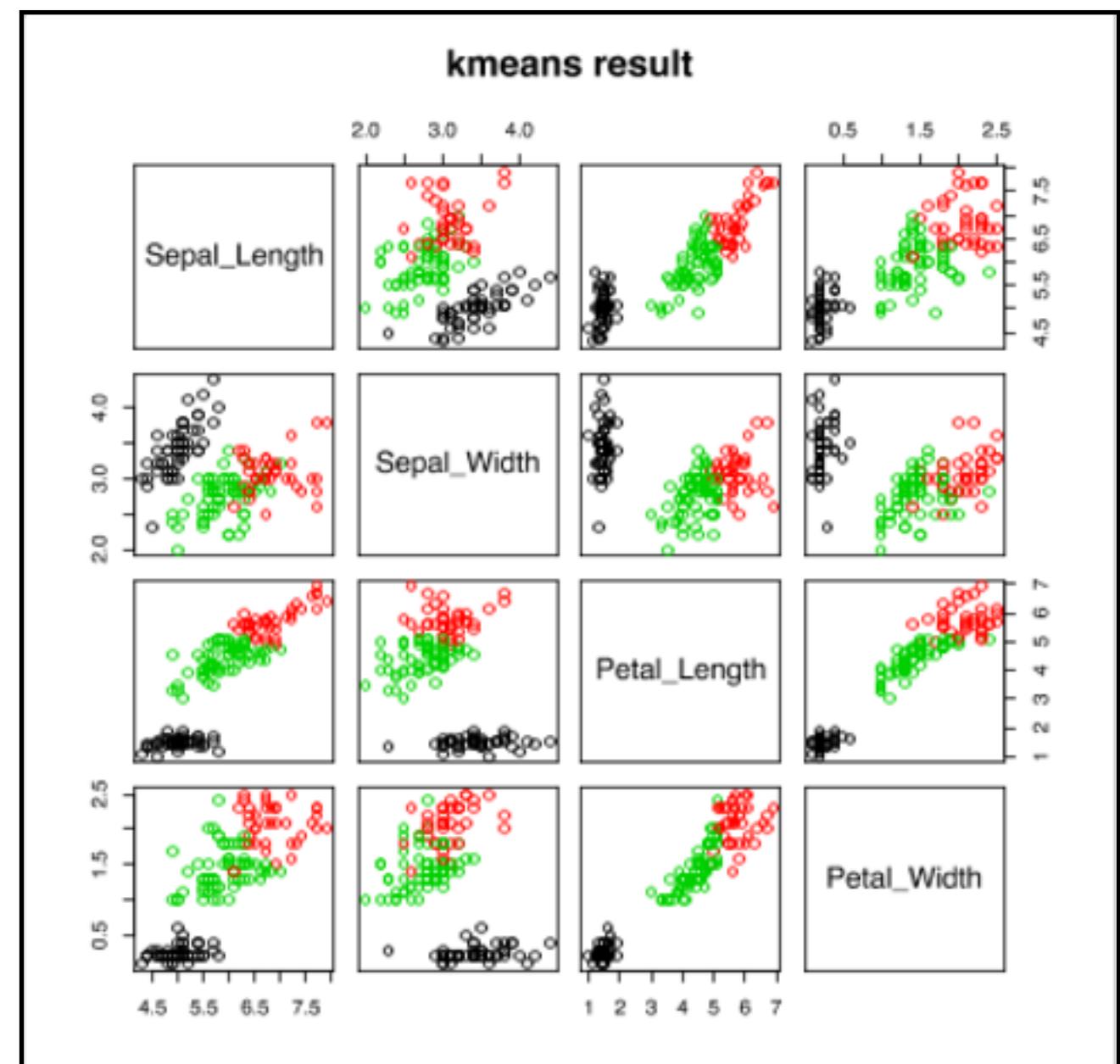
kmeans.json

```
{ "label": "kmeans", "categories": [
  { "variables": {
      "param.columns": { "type": "multi-column-numerical" },
      "param.k": { "type": "int", "default": 3 },
      "param.algorithm": { "items": ["Hartigan-Wong", "Lloyd",
                                    "Forgy", "MacQueen"] },
      "param.plot.title": { "default": "kmeans result" },
      "output.clusterid": { "type": "output-column" }
    }
  }
]}
```

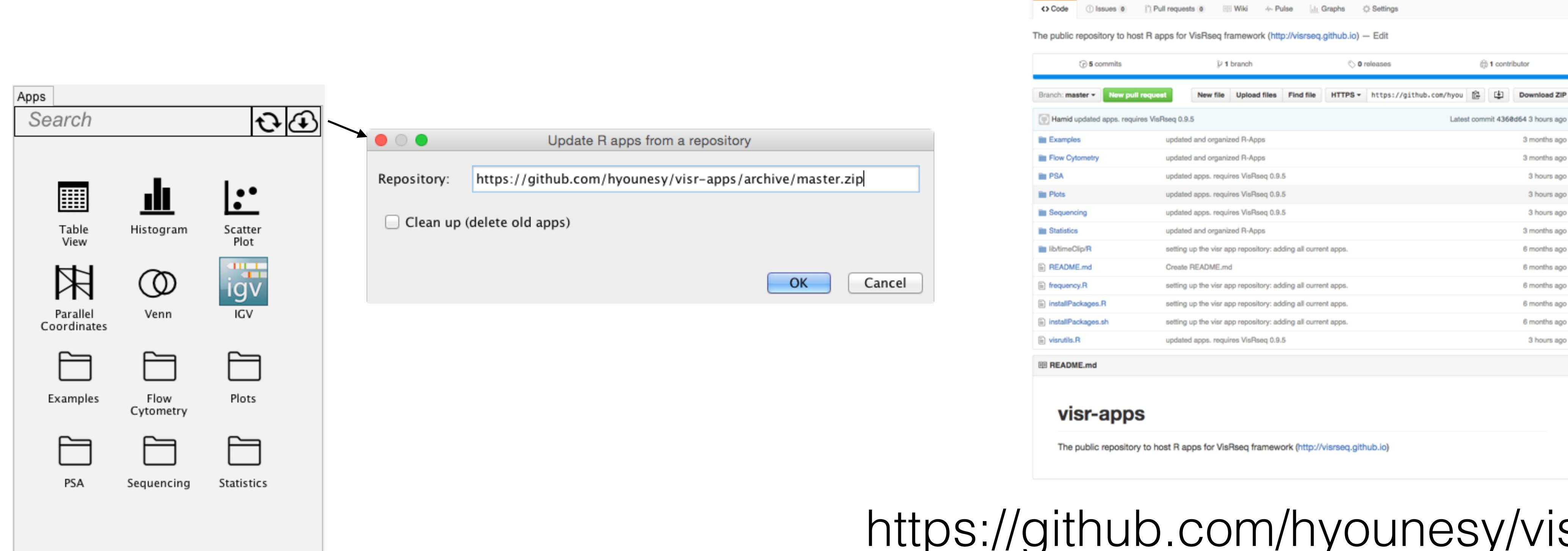
auto generated UI



output



creating R apps: sharing



<https://github.com/hyounesy/visr-apps/>



Assignment

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git

Assignment: Analyzing ML model parameterizations

- How does ML parameterization affects performance?
- Use the full (100,000 row) dataset for analysis
- Write a report discussing what you found
- Single PDF document
- Plots should be generated using only VisRseq
- Other software is ok to add text and arrange/annotate graphics

Assignment 1: Data Analysis

- The submissions will be evaluated based on:
 - functionality / the amount of information content
 - clarity (i.e. ease of interpretation)
 - usability / interaction possibilities
 - interesting findings
- Describe your interesting/surprising findings in words as well.

Assignment 2: App Development

- Develop a new R-App or modify and improve an existing R-App.
- Tutorial: <https://github.com/hyounesy/bioc2016.visrseq>
- Goal is to enhance the current analytical power of VisRseq to allow getting more or improved results.
- App(s) may add new computational functionality (e.g. new classification method) or new plots.
- It is still in the context of movie dataset, but the app should be designed such that it can be used with any tabular data.

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git

Assignment submission

- Due by Thursday, Sept. 7 at 11:59PM CET
- Results will be discussed on Friday morning
- Submission using a **git pull request** of:
 - a directory named your full name (**firstname_lastname**)
 - containing your submissions
 - a readme.txt or readme.md file containing which assignment(s) you are submitting and any additional information about your submission.

acknowledgement

- Developers: Xu Fan, Kathy Cheng, Joseph Poper
- BC Genome Sciences Centre
- Lorincz lab, Underhill lab, Rossi lab, Mager lab



BC Cancer Agency
CARE + RESEARCH



NSERC
CRSNG

<http://bit.ly/2gufWsB>

git clone https://github.com/hyounesy/cass2017_vis.git

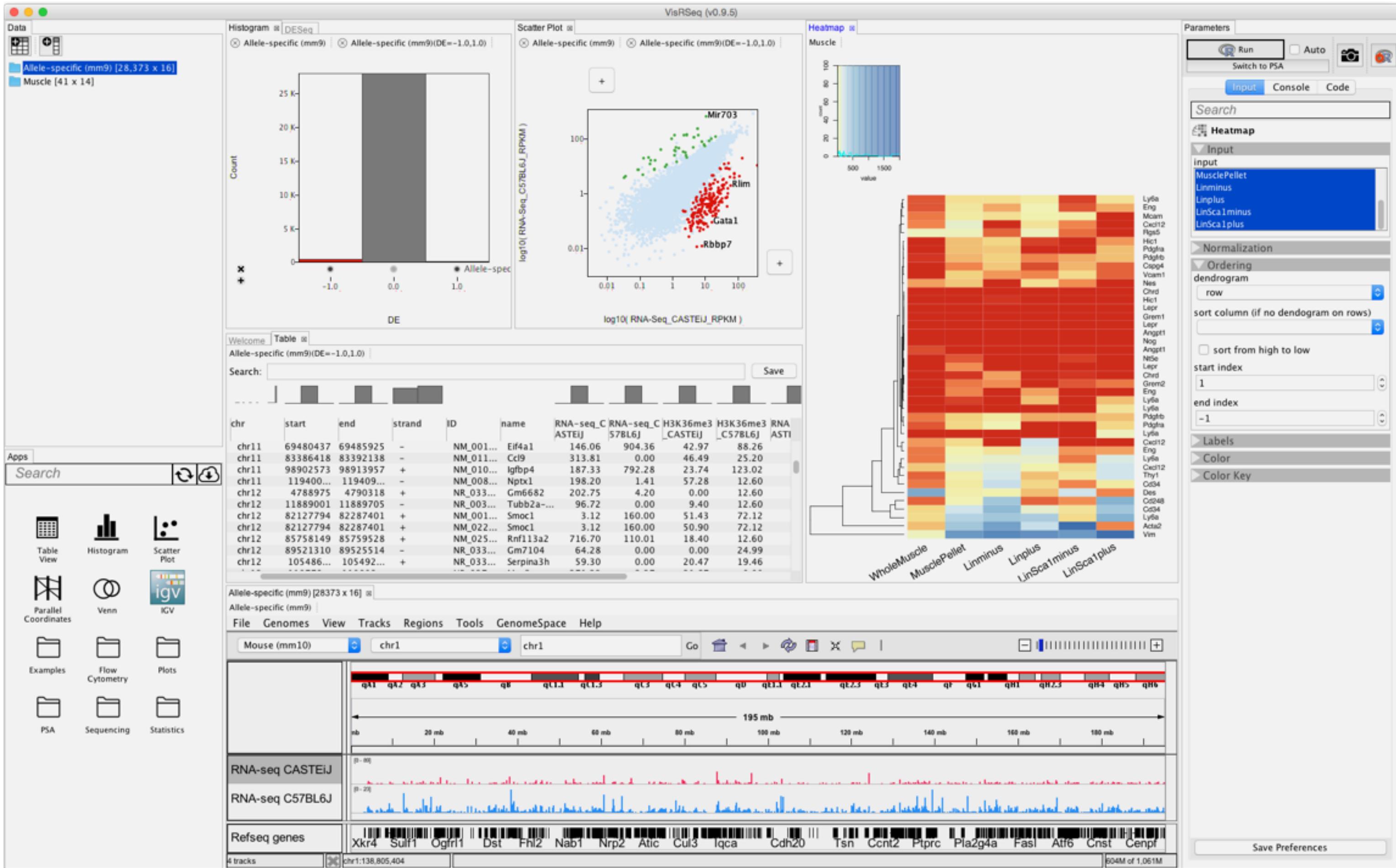
Questions/issues?



- submit an issue on github: https://github.com/hyounesy/cass2017_vis
- email me: thomas.torsney-weir@univie.ac.at

Demo

questions?



visrseq.github.io
(please email for latest version)