

www.leedsomics.org
@leedsomics
omics@leeds.ac.uk

Plotting charts with R

Club Moderators: Elton Vasconcelos, Peter Mulhair, and Dapeng Wang

Topics to be addressed - Survey Result

1st session



R has its own command line environment

Table of Useful R commands

https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

Command	Purpose	Command	Purpose
<code>help()</code>	Obtain documentation for a given R command	<code>plot()</code>	Produces a scatterplot
<code>example()</code>	View some examples on the use of a command	<code>xyplot()</code>	Lattice command for producing a scatterplot
<code>c()</code> , <code>scan()</code>	Enter data manually to a vector in R	<code>lm()</code>	Determine the least-squares regression line
<code>seq()</code>	Make arithmetic progression vector	<code>anova()</code>	Analysis of variance (can use on results of <code>lm()</code>)
<code>rep()</code>	Make vector of repeated values	<code>predict()</code>	Obtain predicted values from linear model
<code>data()</code>	Load (often into a <code>data.frame</code>) built-in dataset	<code>nls()</code>	estimate parameters of a nonlinear model
<code>View()</code>	View dataset in a spreadsheet-type format	<code>residuals()</code>	gives (observed - predicted) for a model fit to data
<code>str()</code>	Display internal structure of an R object	<code>sample()</code>	take a sample from a vector of data
<code>read.csv()</code> , <code>read.table()</code>	Load into a <code>data.frame</code> an existing data file	<code>replicate()</code>	repeat some process a set number of times
<code>library()</code> , <code>require()</code>	Make available an R add-on package	<code>cumsum()</code>	produce running total of values for input vector
<code>dim()</code>	See dimensions (# of rows/cols) of <code>data.frame</code>	<code>ecdf()</code>	builds empirical cumulative distribution function
<code>length()</code>	Give length of a vector	<code>dbinom()</code> , etc.	tools for binomial distributions
<code>ls()</code>	Lists memory contents	<code>dpois()</code> , etc.	tools for Poisson distributions
<code>rm()</code>	Removes an item from memory	<code>pnorm()</code> , etc.	tools for normal distributions
<code>names()</code>	Lists names of variables in a <code>data.frame</code>	<code>qt()</code> , etc.	tools for student <i>t</i> distributions
<code>hist()</code>	Command for producing a histogram	<code>pchisq()</code> , etc.	tools for chi-square distributions
<code>histogram()</code>	Lattice command for producing a histogram	<code>binom.test()</code>	hypothesis test and confidence interval for 1 proportion
<code>stem()</code>	Make a stem plot	<code>prop.test()</code>	inference for 1 proportion using normal approx.
<code>table()</code>	List all values of a variable with frequencies	<code>chisq.test()</code>	carries out a chi-square test
<code>xtabs()</code>	Cross-tabulation tables using formulas	<code>fisher.test()</code>	Fisher test for contingency table
<code>mosaicplot()</code>	Make a mosaic plot	<code>t.test()</code>	student <i>t</i> test for inference on population mean
<code>cut()</code>	Groups values of a variable into larger bins	<code>qqnorm()</code> , <code>qqline()</code>	tools for checking normality
<code>mean()</code> , <code>median()</code>	Identify “center” of distribution	<code>addmargins()</code>	adds marginal sums to an existing table
<code>by()</code>	apply function to a column split by factors	<code>prop.table()</code>	compute proportions from a contingency table
<code>summary()</code>	Display 5-number summary and mean	<code>par()</code>	query and edit graphical settings
<code>var()</code> , <code>sd()</code>	Find variance, sd of values in vector	<code>power.t.test()</code>	power calculations for 1- and 2-sample <i>t</i>
<code>sum()</code>	Add up all values in a vector	<code>anova()</code>	compute analysis of variance table for fitted model
<code>quantile()</code>	Find the position of a quantile in a dataset		
<code>barplot()</code>	Produces a bar graph		
<code>barchart()</code>	Lattice command for producing bar graphs		
<code>boxplot()</code>	Produces a boxplot		
<code>bwplot()</code>	Lattice command for producing boxplots		

https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

http://www.math.umt.edu/olear/stat458/Rseminar_2.pdf



Other useful R material for beginners

Brief explanations on:

Regular **plot** function →

<https://www.datamentor.io/r-programming/plot-function/>

The powerful **ggplot** function →

<http://r-statistics.co/ggplot2-Tutorial-With-R.html>

Brief explanations on:

Regular **plot** function → <https://www.datamentor.io/r-programming/plot-function/>

The powerful **ggplot** function → <http://r-statistics.co/ggplot2-Tutorial-With-R.html>

- A simple “ggplot barchart” code on a biological context

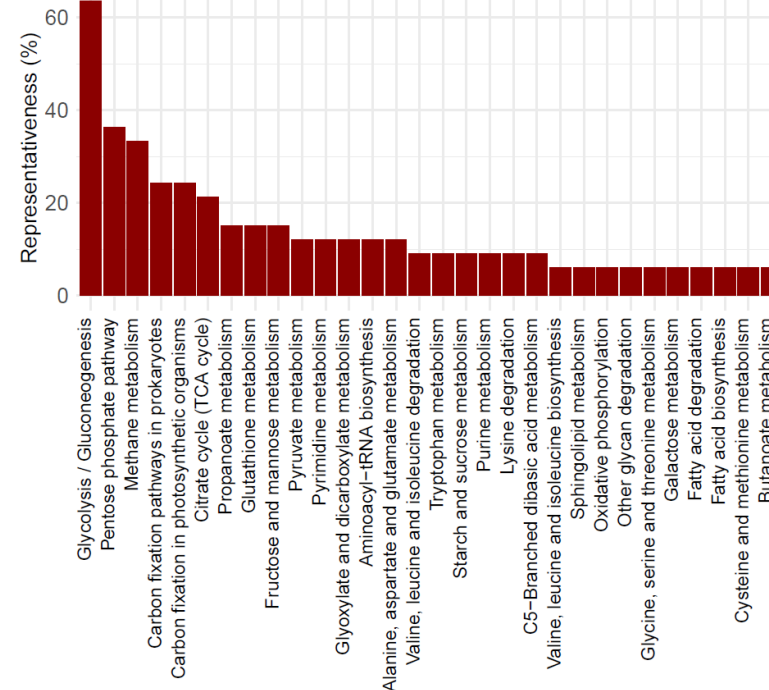
```
### Working on the ones present in at least two or more species (*_gt1.tab)
library(ggplot2)
kegg = read.table("allKOs-counts_gt1.tab", sep="\t")
dim(kegg)
pdf("allKOs-barChart.pdf")
ggplot(data.frame(kegg), aes(x=kegg[,2], y=(kegg[,1]/33)*100)) + geom_bar(stat="identity", fill = "dark red") + theme_minimal() + theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0,size=11.5, colour="black"), axis.text.y=element_text(size=14), axis.title=element_text(size=14), plot.title = element_text(hjust = 0.5, size = 16)) + ggtitle("Significant KEGG pathways") + ylab("Representativeness (%)") + xlab("") + scale_x_discrete(limits=unique(kegg$V2))
dev.off()
```

Input file

```
21 Glycolysis / Gluconeogenesis
12 Pentose phosphate pathway
11 Methane metabolism
8 Carbon fixation pathways in prokaryotes
8 Carbon fixation in photosynthetic organisms
7 Citrate cycle (TCA cycle)
5 Propanoate metabolism
5 Glutathione metabolism
5 Fructose and mannose metabolism
4 Pyruvate metabolism
4 Pyrimidine metabolism
4 Glyoxylate and dicarboxylate metabolism
4 Aminoacyl-tRNA biosynthesis
4 Alanine, aspartate and glutamate metabolism
3 Valine, leucine and isoleucine degradation
3 Tryptophan metabolism
3 Starch and sucrose metabolism
3 Purine metabolism
3 Lysine degradation
3 C5-Branched dibasic acid metabolism
2 Valine, leucine and isoleucine biosynthesis
2 Sphingolipid metabolism
2 Oxidative phosphorylation
2 Other glycan degradation
2 Glycine, serine and threonine metabolism
2 Galactose metabolism
2 Fatty acid degradation
2 Fatty acid biosynthesis
2 Cysteine and methionine metabolism
2 Butanoate metabolism
allKOs-counts_gt1.tab (END)
```



Significant KEGG pathways



Some important ggplot functions and plot types

Function	Chart type
geom_bar	Bar chart
geom_point	Scatter plot
geom_line	Line chart
geom_jitter	Stripchart
geom_boxplot	Boxplot
geom_violin	Violin plot
geom_tile	Heatmap

Bring your issues on!



Registrations open for the

Introduction to RNA-Seq Analysis WORKSHOP

26-27th June, 2019

Worsley Bld., level 11, 11.06

More information at www.leedsomics.org