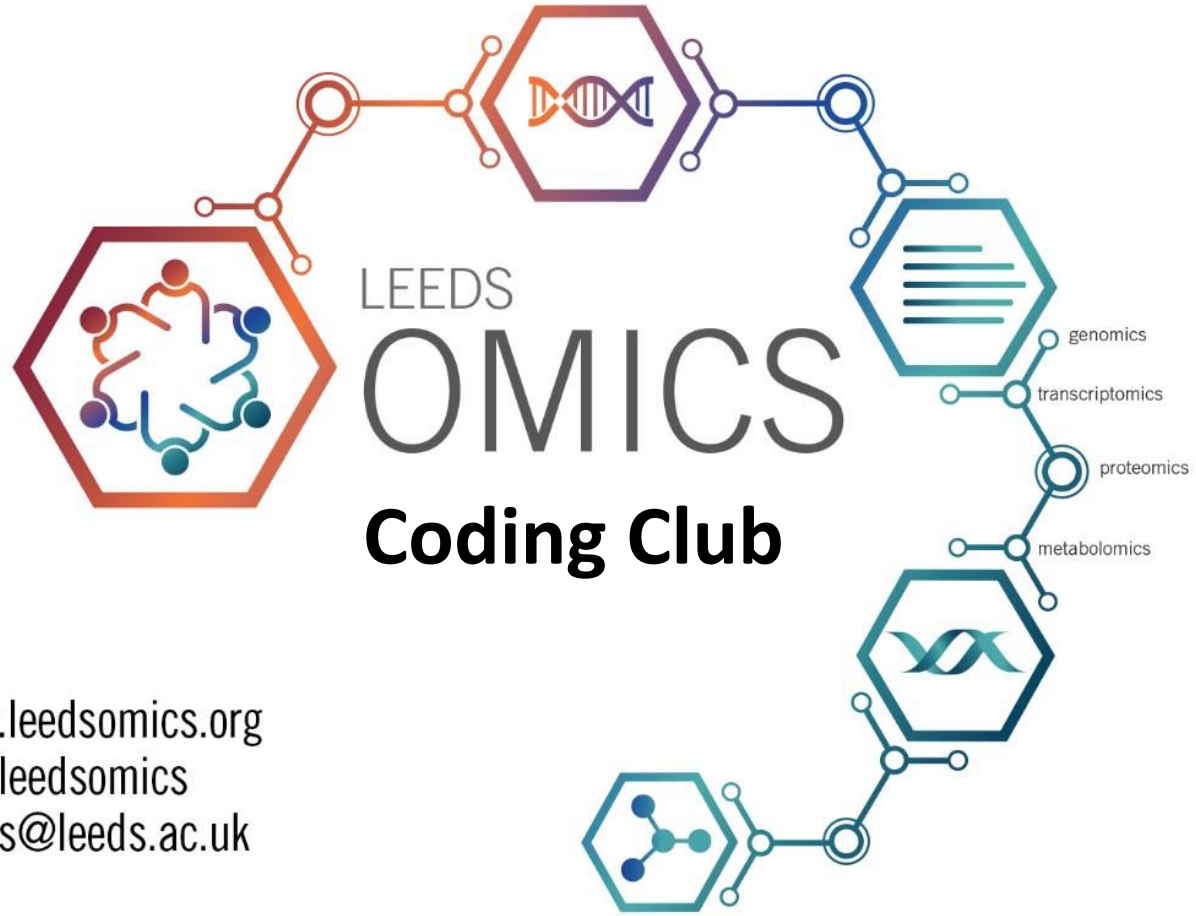


Coding Club

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

Plotting charts with ggplot2 in R

Club Moderators: Elton Vasconcelos, Euan McDonell

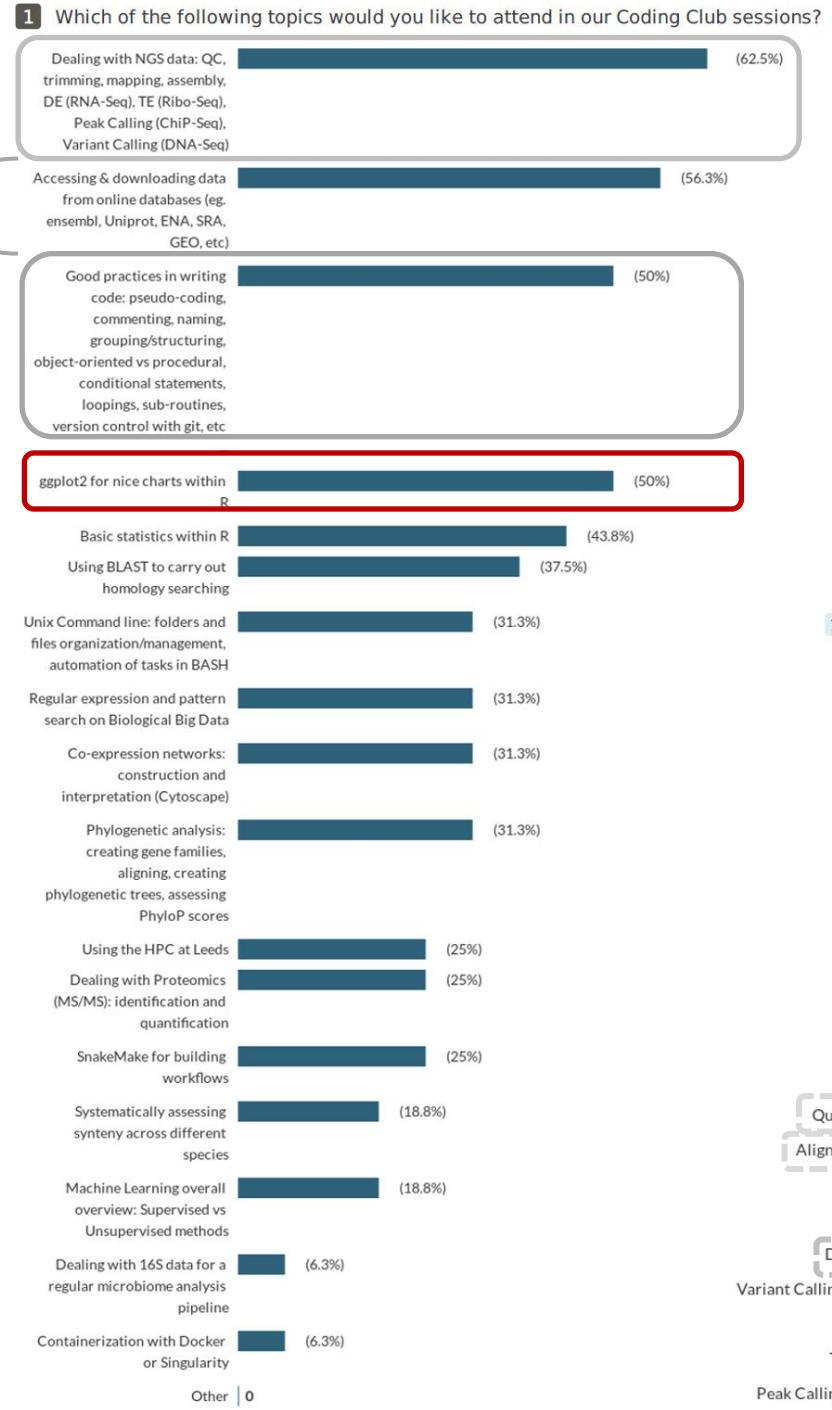
Topics to be addressed on the **2020-21** season - **Survey Result**

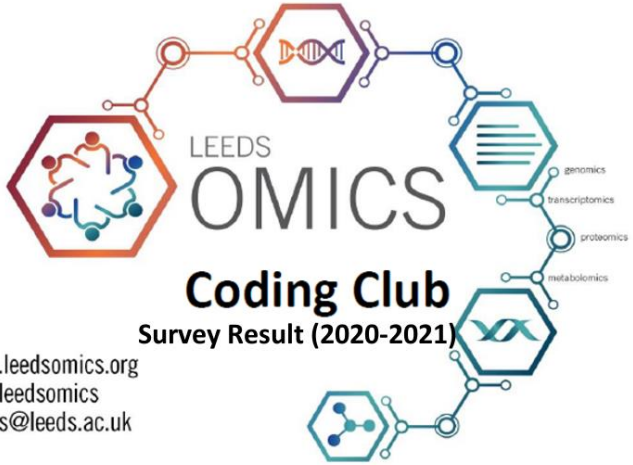
1st, 4th, 6th sessions

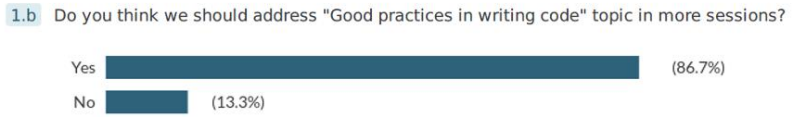
2nd session

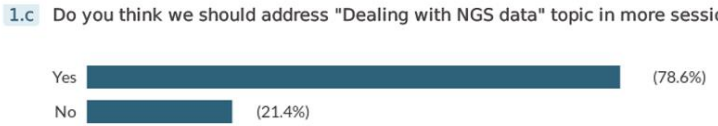
3rd, 5th, 7th sessions

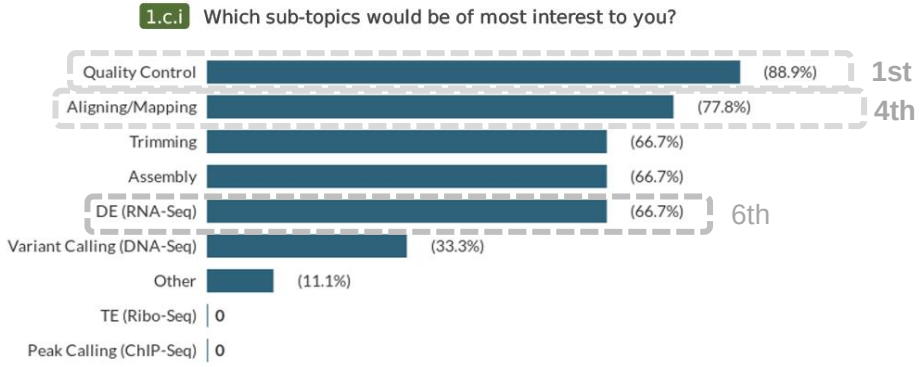
8th session











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
help()	Obtain documentation for a given R command
example()	View some examples on the use of a command
c(), scan()	Enter data manually to a vector in R
seq()	Make arithmetic progression vector
rep()	Make vector of repeated values
data()	Load (often into a data.frame) built-in dataset
View()	View dataset in a spreadsheet-type format
str()	Display internal structure of an R object
read.csv(), read.table()	Load into a data.frame an existing data file
library(), require()	Make available an R add-on package
dim()	See dimensions (# of rows/cols) of data.frame
length()	Give length of a vector
ls()	Lists memory contents
rm()	Removes an item from memory
names()	Lists names of variables in a data.frame
hist()	Command for producing a histogram
histogram()	Lattice command for producing a histogram
stem()	Make a stem plot
table()	List all values of a variable with frequencies
xtabs()	Cross-tabulation tables using formulas
mosaicplot()	Make a mosaic plot
cut()	Groups values of a variable into larger bins
mean(), median()	Identify "center" of distribution
by()	apply function to a column split by factors
summary()	Display 5-number summary and mean
var(), sd()	Find variance, sd of values in vector
sum()	Add up all values in a vector
quantile()	Find the position of a quantile in a dataset
barplot()	Produces a bar graph
barchart()	Lattice command for producing bar graphs
boxplot()	Produces a boxplot
bwplot()	Lattice command for producing boxplots

Command	Purpose
plot()	Produces a scatterplot
xyplot()	Lattice command for producing a scatterplot
lm()	Determine the least-squares regression line
anova()	Analysis of variance (can use on results of lm())
predict()	Obtain predicted values from linear model
nls()	estimate parameters of a nonlinear model
residuals()	gives (observed - predicted) for a model fit to data
sample()	take a sample from a vector of data
replicate()	repeat some process a set number of times
cumsum()	produce running total of values for input vector
ecdf()	builds empirical cumulative distribution function
dbinom(), etc.	tools for binomial distributions
dpois(), etc.	tools for Poisson distributions
pnorm(), etc.	tools for normal distributions
qt(), etc.	tools for student t distributions
pchisq(), etc.	tools for chi-square distributions
binom.test()	hypothesis test and confidence interval for 1 proportion
prop.test()	inference for 1 proportion using normal approx.
chisq.test()	carries out a chi-square test
fisher.test()	Fisher test for contingency table
t.test()	student t test for inference on population mean
qqnorm(), qqline()	tools for checking normality
addmargins()	adds marginal sums to an existing table
prop.table()	compute proportions from a contingency table
par()	query and edit graphical settings
power.t.test()	power calculations for 1- and 2-sample t
anova()	compute analysis of variance table for fitted model

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 **ggplot2** function → http://r-statistics.co/ggplot2-Tutorial-With-R.html

- install.packages("ggplot2")
- library(ggplot2)

Some important ggplot functions and plot types

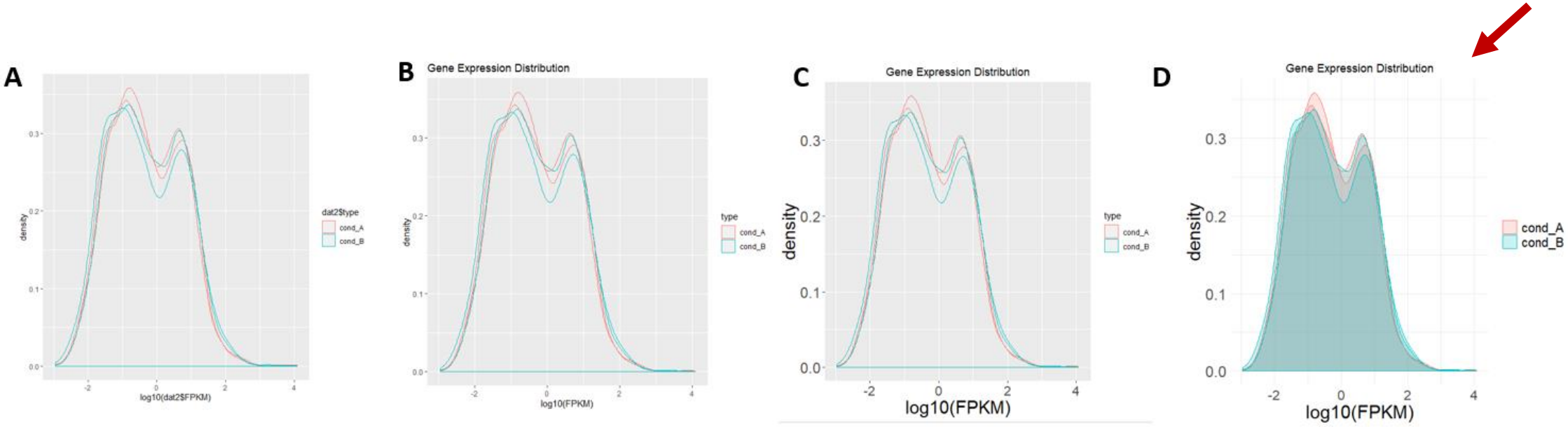
Function	Description
ggplot()	Create a new ggplot using a data frame as input
aes()	Construct aesthetic mappings (goes within ggplot brackets)
+	Add components to a plot
geom_	Geometric object (the actual chart type)
theme()	Set x and y axes parameters
ggtitle()	Set a title
ggsave()	Save a ggplot (or other grid object) with sensible defaults

Sub-function	Chart type
geom_density	Histogram
geom_boxplot	Boxplot
geom_jitter	Stripchart
geom_violin	Violin plot
geom_point	Scatter plot
geom_bar	Bar and pie charts
geom_tile	Heatmaps

FPKM histogram as an example:

```
> dat2 <- data.frame(samples = sampNames, FPKM = expVals, type = conditions, geneID = rownames(dat))
```

```
> ggplot(dat2, aes(x = log10(FPKM), element_line = samples, color = type, fill = type)) +
geom_density(alpha=0.2) + theme_minimal() + ggtitle("Gene Expression Distribution") + theme(plot.title =
element_text(hjust=0.5), axis.text.x = element_text(size=16), axis.text.y=element_text(size=18),
axis.title=element_text(size=22), legend.title = element_blank(), legend.text = element_text(size = 15))
```



Bring your issues on!