

Introducing R

Dr Sonika Tyagi

Monash Bioinformatics Platform Manager
Monash Data Fluency



@MonashBioinfo

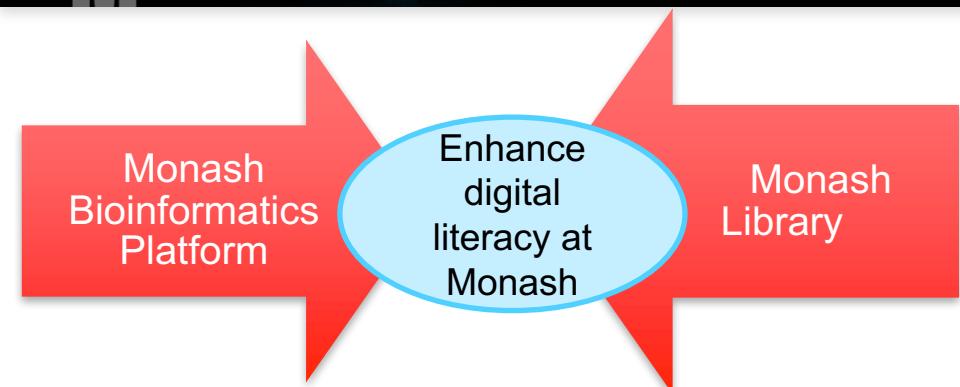


Bioinformatics.platform@monash.edu



<http://bioinformatics.erc.monash.edu>

Monash Data Fluency officially launched on 23.03.2018



MONASH University

my.monash Current students Staff Library Giving to Monash

DATA FLUENCY

HOME ABOUT WORKSHOPS COMMUNITY OF PRACTICE UPCOMING EVENTS

MONASH DATA FLUENCY FOR RESEARCH

Learn more

BECOME DATA FLUENT

What does it mean to be data fluent? For researchers, this means knowing how to use, explore, interpret and visualise data in a meaningful way and effectively communicate your research and ideas. There is an increasing need for all researchers and professionals to be able to understand and write code and interact in the digital enquiry space. That's what data fluency enables you to do.

Community of Practice

Learn, share and enhance your data fluency skills.

Find out more

Workshops

Data fluency skill development opportunities.

Find out more

Upcoming events

03 Aug Introduction to HPC 3 August 2018

Keep up to date

Would you like to join our mailing list and keep up to date with Data Fluency workshops, tech talks and more?

Press the Join now button below and sign up for our mailing list.

If you have an enquiry, you're welcome to email us at datafluency@monash.edu

Get involved

Find out how you can get involved

Find out more

Learning resources

Engage with our learning resources

Find out more

Join now

<https://www.monash.edu/data-fluency>

Monash Data Fluency activities



15 Carpentry trained instructors and counting

Building a network
of instructors and
learners across the
Monash campuses

Workshops:

4 at Clayton campus

2 at the Peninsula campus

1 running today.

1 on 3 Aug 2018 at Clayton

Friday drop in session:

Every Friday 3pm at G19, 15 innovation walk

Special monthly Fri drop-ins (start 2pm)

R Book club

Get involved:

Email: datafluency@monash.edu

Follow on twitter: [@resdatflu](https://twitter.com/@resdatflu)

Slack <https://datafluency.slack.com>

<https://www.monash.edu/data-fluency>

Today's Objective(s) – What we hope you will get out of today

- Introduction to R
- The basics of R – data objects
- Data plotting
- How to find more information for a specific task
- The start of the journey, a network of R learners/users around you.

Trainers – Paul Harrison, Adele Barugahare, Sarah Williams

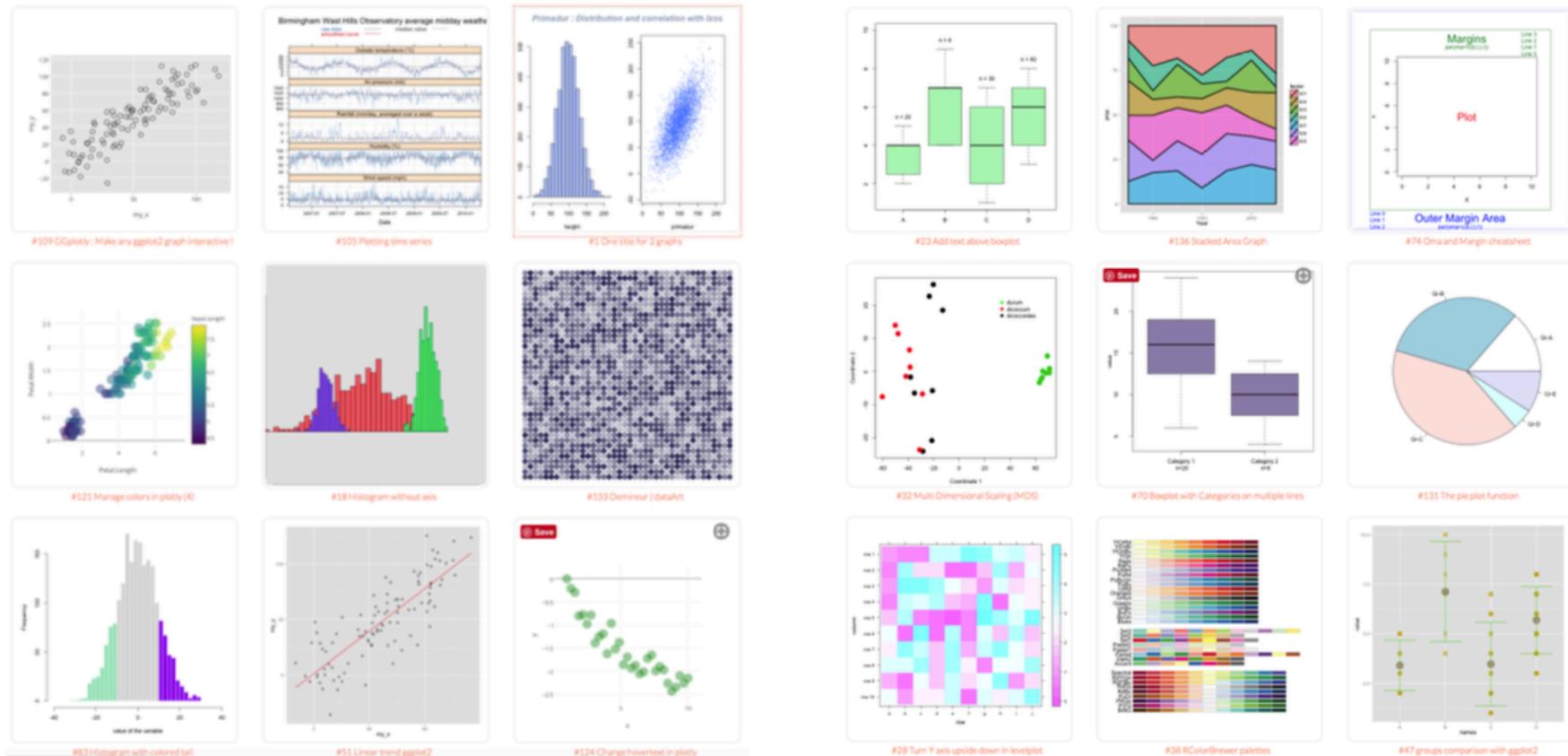
Workshop Info: <https://tinyurl.com/R20Jul18>

Why R? (as opposed to Excel?)

- It's free, legally free. Active development.
- Reproducible analysis.
 - Document what you have done with your data in code.
 - Come back to it in days, months, years and you should know what was done.
- Low risk of inadvertent data loss/mutation.
 - By design, R requires you to load your data in, what you do with the data is then written in code (R language).
- R can handle really large datasets.
 - Excel is limited by 1,048,576 rows and 16,384 columns.
- Collaborative.
 - Share your data and analysis.

Why R? (as opposed to Excel?)

- Graphs and plotting.
- Can you plot a box and whisker plot in excel?



How to get R?

- Many specialised tools/libraries for specific purposes
 - Collection of functions

<https://cran.r-project.org>)



[CRAN](#)
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

[About R](#)
[R Homepage](#)
[The R Journal](#)

[Software](#)
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

[Documentation](#)
[Manuals](#)
[FAQs](#)
[Contributed](#)

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Thursday 2017-09-28, Short Summer) [R-3.4.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

R and the interfaces to R

▪ R on the command line

```
nwon0008 — R — 80x24
Last login: Thu Oct  5 11:06:23 on ttys001
[MU00105304X:~ nwon0008$ R]

R version 3.4.1 (2017-06-30) -- "Single Candle"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> ]
```

R and the interfaces to R

▪ RStudio – RStudio Server, a graphical interface to R

The screenshot shows the RStudio interface. On the left, the Console tab displays the standard R startup message. On the right, the Environment tab shows R code related to the Iris dataset, specifically setting ranges for Petal Width and Petal Length. Below the code, the Viewer tab displays a page titled "Linear Models for Microarray Data" from the LIMMA package, featuring the R logo and navigation links for user guides and package vignettes.

```
R version 3.4.1 (2017-06-30) -- "Single Candle"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

list(range = c(0,2.5),
label = 'Petal Width', values = ~petal_width),
list(range = c(1,7),
label = 'Petal Length', values = ~petal_length)
)
)
p
q("no")
q("no")
q("no")

Linear Models for Microarray Data

User Guides and Package Vignettes

- [LIMMA User's Guide \(pdf\)](#). This is the main documentation for the package.
- [LIMMA Introduction \(pdf\)](#). One page introduction.
- [LIMMA Change Log \(txt\)](#). Historical record of changes.

[Package Contents]

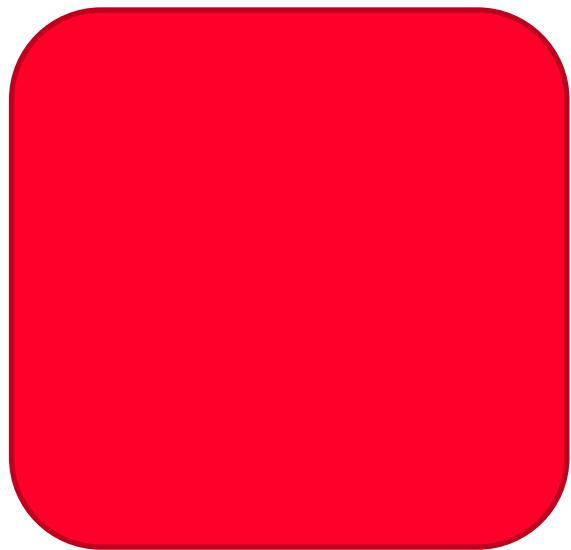
Today's agenda

- Starting out in R
- Working with data in a matrix
- Working with data frames
- Plotting with ggplot2

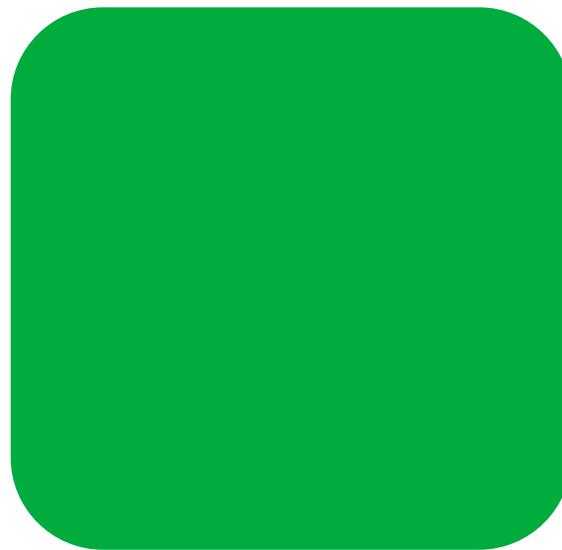


How today's workshop works – signalling for help

- Workshop notes
- Sticky notes



I need help please



I am all good

Working with data in a matrix

Explanatory text

Loading data

Our example data is quality measurements (particle size) on PVC plastic production, using eight different resin batches, and three different machine operators.

The data set is stored in comma-separated value (CSV) format. Each row is a resin batch, and each column is an operator. In RStudio, open pvc.csv and have a look at what it contains.

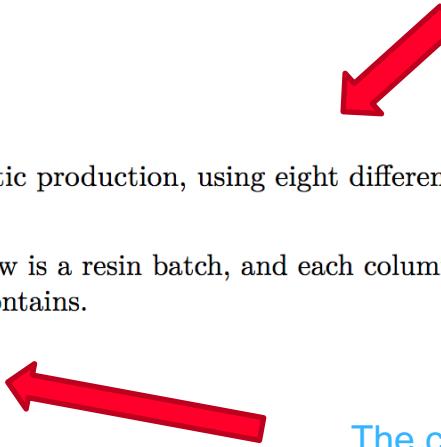
```
read.csv("r-intro-files/pvc.csv", row.names=1)
```

Tip

The location of the file is given relative to your “working directory”. You can see the location of your working directory in the title of the console pane in RStudio. It is most likely “~”, indicating your personal home directory. You can change working directory with `setwd`.

The filename “r-intro-files/pvc.csv” means from the current working directory, in the sub-directory “r-intro-files”, the file “pvc.csv”.

You can check that the file is actually in this location using the “Files” pane in the bottom right corner of RStudio.



The code

The R workbook – Challenges and extra homework

```
avg_operator <- apply(mat, 2, mean)
```

Since the second argument to `apply` is MARGIN, the above command is equivalent to `apply(dat, MARGIN = 2, mean)`.

Tip

Some common operations have more concise alternatives. For example, you can calculate the row-wise or column-wise means with `rowMeans` and `colMeans`, respectively.

Challenge - summarizing the matrix

How would you calculate the standard deviation for each resin?

Advanced: How would you calculate the values two standard deviations above and below the mean for each resin?



We will give you challenges through the workshop to work with the example data, also homework

Rounding up

- This is the start of your R journey, many others are at the same stage, share your questions.
- Attend this workshop again, and others we offer
 - Advanced R
 - Specialised case studies using R
 - R book club
- MBP have Friday help sessions at Clayton 3:00pm.
- Monthly special theme Friday drop-in 2pm onwards
 - Subscribe to mailing list to get a notification (look at data fluency website)

Ice breaker activity

- Introduce yourself to your neighbour (s) and tell them why you are here and what you want to achieve.



Thank you !

- **Paul Harrison**
- **Sarah Williams**
- **Adele Barugahare**
- Andrew Perry
- Nick Wong

- Sebastian Borutta
- Gunta Jaudzems
- Debbie Symons
- Linda Kalejs

- David Powell
- David Groenewegen

- The MBP and the library team
- Data Fluency CoP

Workshop Info: <https://tinyurl.com/R20Jul18>