To share your e-Eso experience use:
#e_ESO

Introduction to statistical analysis with R software for cancer scientists

Expert: **Prof Livio Fenga**, Senior Lecturer in Analytics, University of Exeter, Exeter, UK

Discussant: **Dr Luca Bertolaccini**, IEO European Institute of Oncology IRCCS, Milan, Italy

Introduction to statistical analysis with R software for Cancer Scientists

Livio Fenga: Senior Lecturer in Analytics, University of Exeter (UK)

Webinar e–ESO, June 16, 2022

Outline

1. R overview
2. Mastering R…no thanks. Let's use it
3. *Changing paradigm*: From plots to LATTICES

   (Tumor growth is inherently a three-dimensional process).
5. How can R help us (PRACTICALLY)

5A **Neural Network with cancer data with R**

5B **R and the power of simulation for cancer data**

(**Creating and visualizing spatial simulations of tumor growth**)

R overview but, before we start, some Q&As

As an MD ***and*** Researcher do I need to know R?

YES

As an MD ***and*** Researcher can I learn R?

YES

How?

Tutorials AND other *ad hoc* resources;
learning by doing (COPY AND PASTE!)
interdisciplinary links among different depts.;

As an MD ***and*** Researcher should I know Statistics?

Being a good practitioner is usually enough
Again **interdisciplinary links (statisticians, mathematicians, physicists) are vital**

Download R and RStudio

- Download R :

  http://cran.r-project.org/bin/

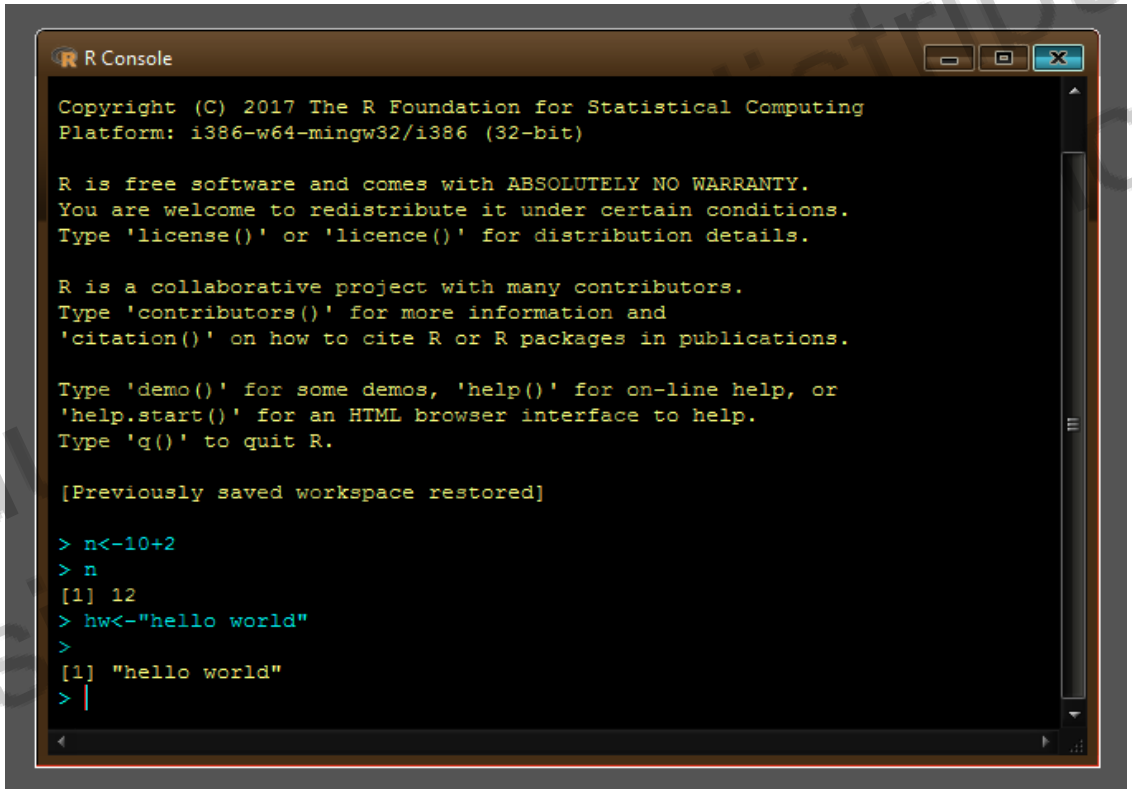
Installation

Installing R on windows PC :

- Use internet browser to point to : http://mirror.aarnet.edu.au/pub/CRAN
- Under the heading Precompiled Binary Distributions, choose the link Windows.
- Next heading is R for Windows; choose the link base.
- Click on download option(R xxxxx for windows).
- Save this to the folder C:\R on your PC.
- When downloading is complete, close or minimize the Internet browser.
- Double click on R 3.4.1-win32.exe in C:\R to install.

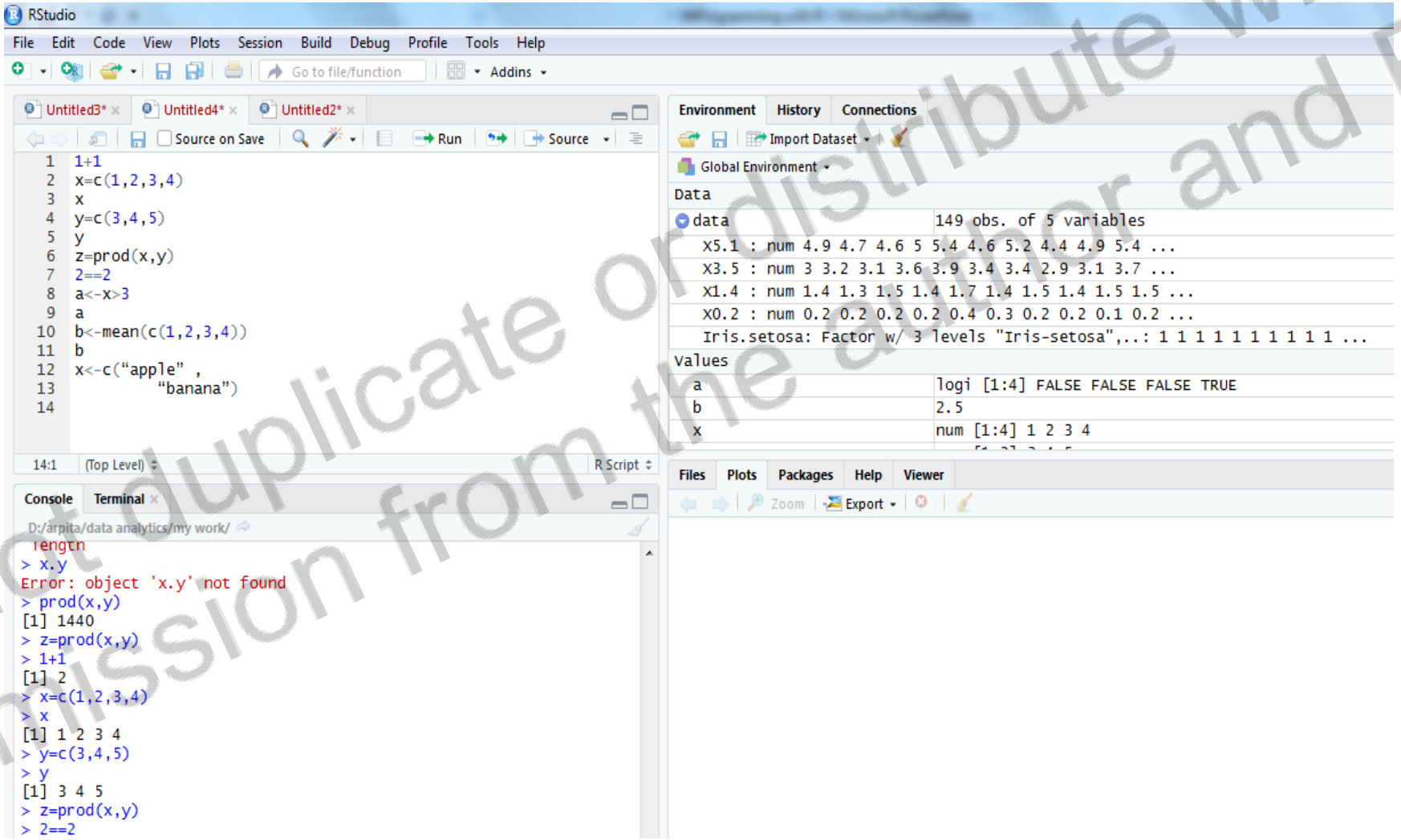
R is Relatively Easy

- Double click the R icon on the Desktop and the R Console will open.
- Wait while the program loads. You observe something like this.



- You can type your own program at the prompt line >.

R command in integrated environment



How to use R for simple maths

- > 3+5
- > 12 + 3 / 4 – 5 + 3*8
- > (12 + 3 / 4 – 5) + 3*8
- > pi * 2^3 – sqrt(4)
- >factorial(4)
- >log(2,10)
- >log(2, base=10)
- >log10(2)
- >log(2)

Note

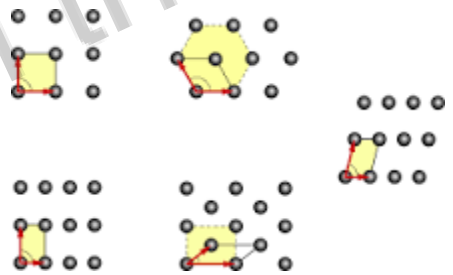
- **R ignores spaces**

Lattice

Lattice data are **observations from a random process observed**

over a countable collection of spatial regions, and supplemented by a neighborhood structure.

The observation locations can be regular (equally spaced grid) or irregular, and data at a particular location typically represent the entire region.



a regular geometrical arrangement of points or objects over an area or in space*specifically* **:** the arrangement of atoms in a crystal

HOW R PRACTICALLY HELP US

canceR

platforms all | rank 1021 / 2140 | support 1 / 4 | in Bioc 7 years
build ok | updated < 3 months | dependencies 155

DOI: 10.18129/B9.bioc.canceR

A Graphical User Interface for accessing and modeling the Cancer Genomics Data of MSKCC

Bioconductor version: Release (3.15)

The package is user friendly interface based on the cgdsr and other modeling packages to explore, compare, and analyse all available Cancer Data (Clinical data, Gene Mutation, Gene Methylation, Gene Expression, Protein Phosphorylation, Copy Number Alteration) hosted by the Computational Biology Center at Memorial-Sloan-Kettering Cancer Center (MSKCC).

Author: Karim Mezhoud. Nuclear Safety & Security Department. Nuclear Science Center of Tunisia.

Maintainer: Karim Mezhoud <kmezhoud at gmail.com>

Citation (from within R, enter `citation("canceR")`):

Tunisia. KMNS&SDNSCo (2022), *canceR: A Graphical User Interface for accessing and modeling the Cancer Genomics Data of MSKCC*. R package version 1.30.01.

Installation

To install this package, start R (version "4.2") and enter:

```
if (!require("BiocManager", quietly = TRUE))
    install.packages("BiocManager")

BiocManager::install("canceR")
```

For older versions of R, please refer to the appropriate Bioconductor release.

Documentation

To view documentation for the version of this package installed in your system, start R and enter:

```
browseVignettes("canceR")
```

Documentation »

Bioconductor

- Package vignettes and manuals.
- Workflows for learning and use.
- Several online books for comprehensive coverage of a particular research field, biological question, or technology.
- Course and conference material.
- Videos.
- Community resources and tutorials.

R / CRAN packages and documentation

Support »

Please read the posting guide. Post questions about Bioconductor to one of the following locations:

- Support site - for questions about Bioconductor packages
- Bioc-devel mailing list - for package developers

1. Neural network for cancer research

2. simulations of tumor growth

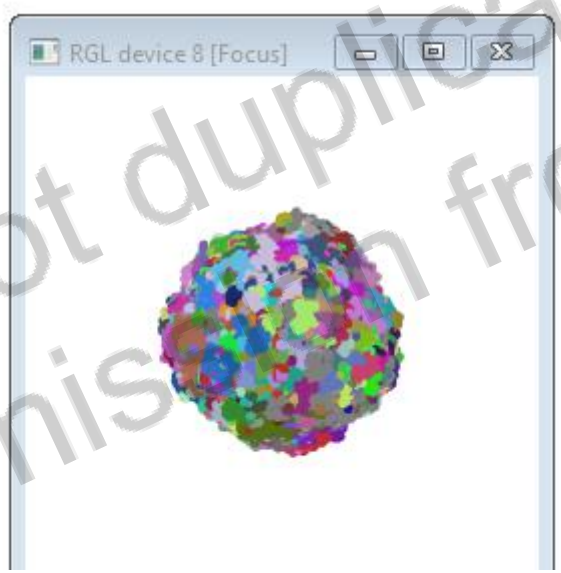
Practical sessions – share screen mode

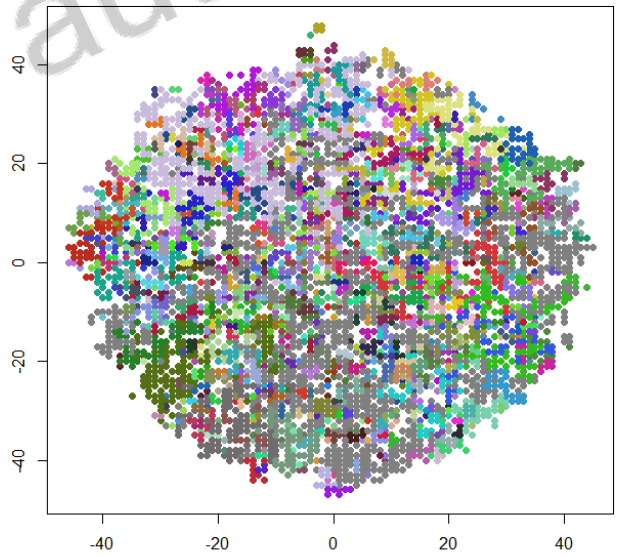
Creating and visualizing spatial simulations of tumor growth using SITH

The cells within a cancerous tumor are usually highly diverse. An average tumor contains hundreds of thousands of mutations spread throughout billions of cancer cells, although it is thought that only a small percentage of these mutations are “drivers” which facilitate the progression of cancer into later stages (Greaves and Maley 2012). A lack of understanding about the **evolutionary** process which results in the observed intratumor heterogeneity is a major obstacle preventing the development of effective cancer therapies (Stanta and Bonin 2018).

```
> out <- simulateTumor(max_pop = 250000, verbose = FALSE)
> visualizeTumor(out, background = "white")
>
> names(out)
[1] "cell_ids"     "genotypes"    "muts"         "phylo_tree"   "color_scheme" "drivers"
[7] "time"         "params"
> head(out$cell_ids)
    x   y  z genotype nmuts distance
1  15 -16 17     6358     2 27.74887
2   1 -23 32       16     2 39.42081
3 -11  19 28     1685     3 35.58089
4  13  -5 34    27691     2 36.74235
5  32  17 -3    64270     7 36.35932
6  43   1 -9    34824     4 43.94315
> head(out$muts)        ## mutation allele frequency (MAF)
  id  count MAF
1  0 250000   1
2  1      0   0
3  2      0   0
4  3      0   0
5  4      0   0
6  5      0   0
>
>
> visualizeTumor(out, background = "white")                    ##  3D plot of the simulated tumor
> visualizeTumor(out, background = "white")
>
>
> par(mfrow = c(1,2))
> plotSlice(tumor = out)
> plotSlice(tumor = out, plot.type = "heat")
> plotSlice(tumor = out)
> visualizeTumor(out, background = "white")
> sp <- spatialDistribution(tumor = out)
```

```
library( STIM )

set.seed(1126490984)

out <- simulateTumor(max_pop = 250000, verbose = FALSE)
visualizeTumor(out, background = "white")

names(out)
head(out$cell_ids)
head(out$muts)        ## mutation allele frequency (MAF)


visualizeTumor(out, background = "white")                    ##  3D plot of the simulated tumor

visualizeTumor(out, background = "white", plot.type = "heat")   ## which colors cells on a scale from blue to red,
                                                                ## depending on the number of mutations within the cell.


par(mfrow = c(1,2))
plotSlice(tumor = out)
plotSlice(tumor = out, plot.type = "heat")


                                                                # Spatial distribution of mutants

sp <- spatialDistribution(tumor = out)
```





Take home concepts

Use R knowing that it is not your core business

Interdisciplinary approach!
Alone we are the
"nothing" proceeding in the
"Nihil"

Use grants, collaborations, students exchanges to establish a statistical R-based community within your research groups

R is free and the information/help
Available on the net can solve most of the issues

Questions? Comments?

THANK YOU!!