

2- Data Visualization

Curso Ciencia de datos en R

Escuela de Doctorado, Universidad de Alcalá, Mayo-Junio 2021

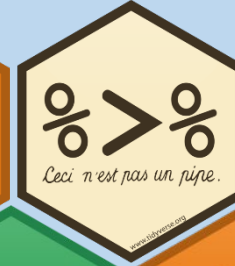
Prof. Sara Villén Pérez

0 – Intro to R



R Studio®

1 - Data manipulation



2 – Data visualization

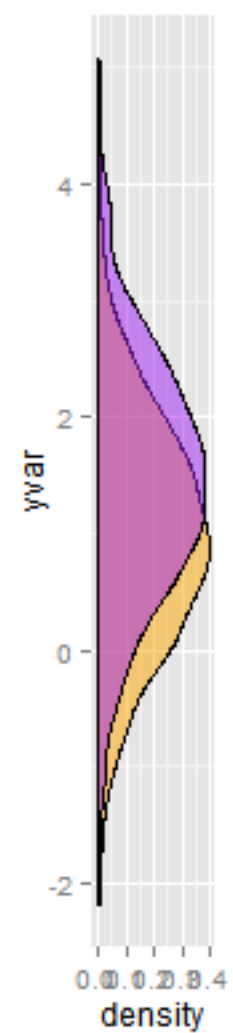
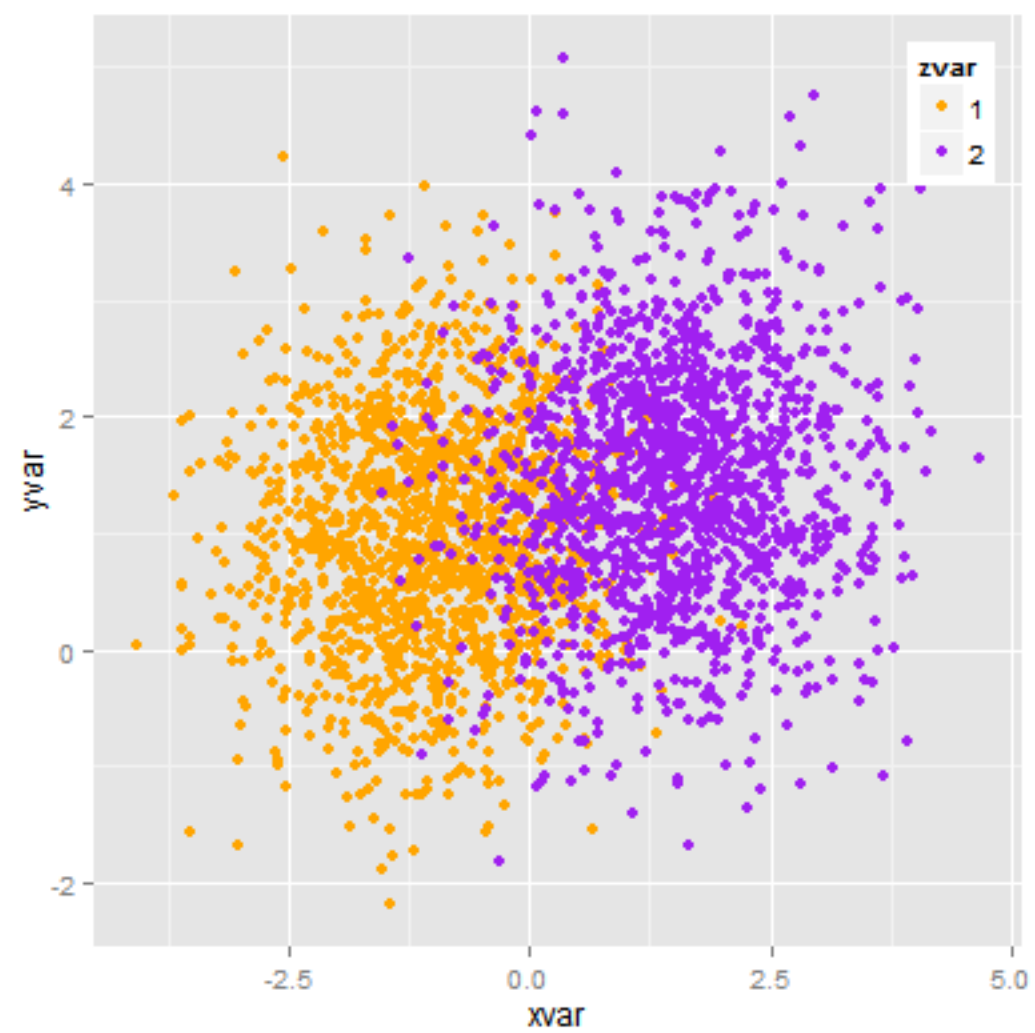
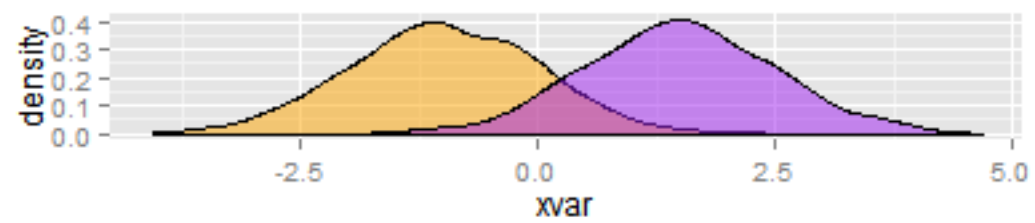


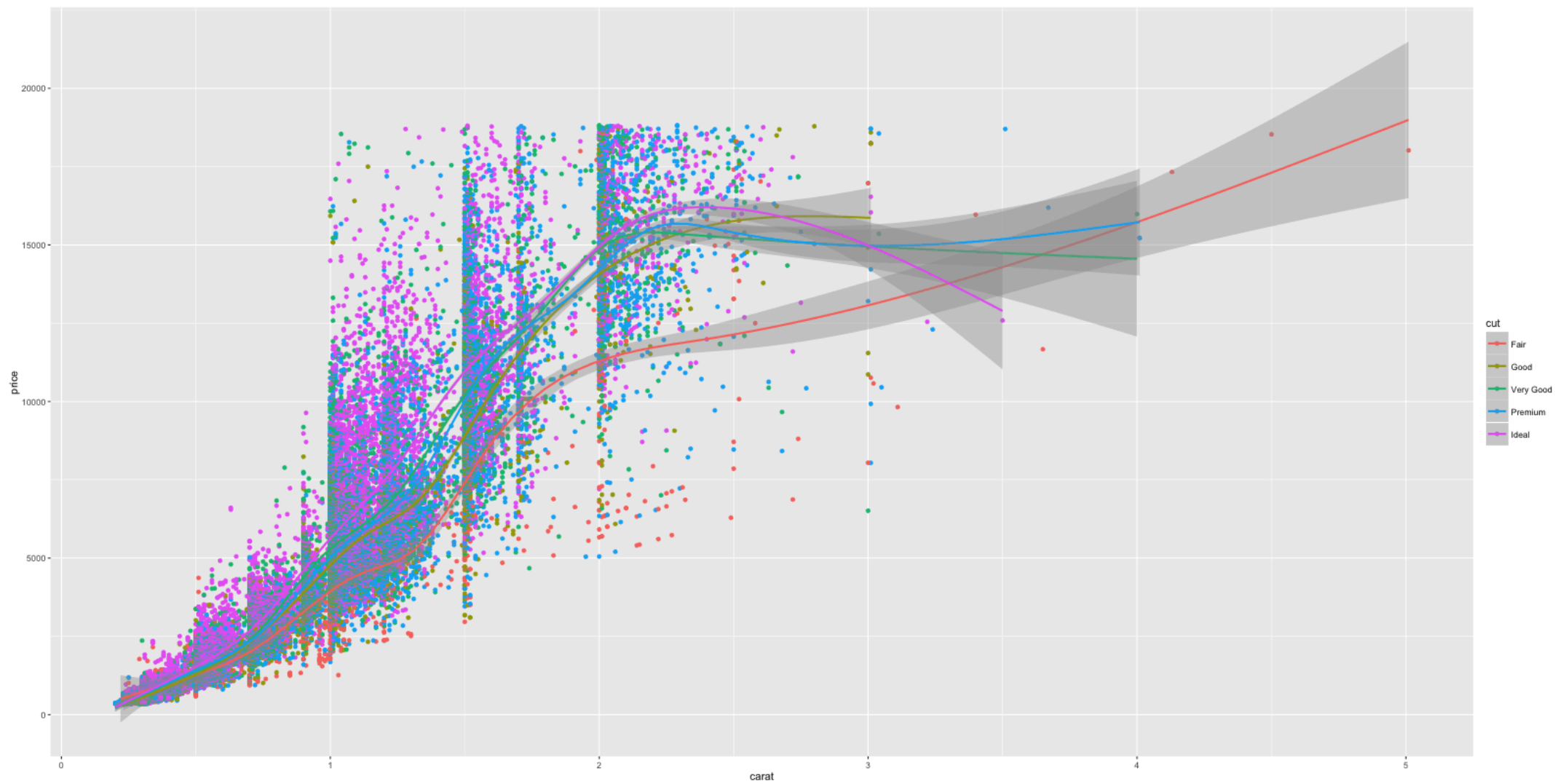
3 – Automation

for

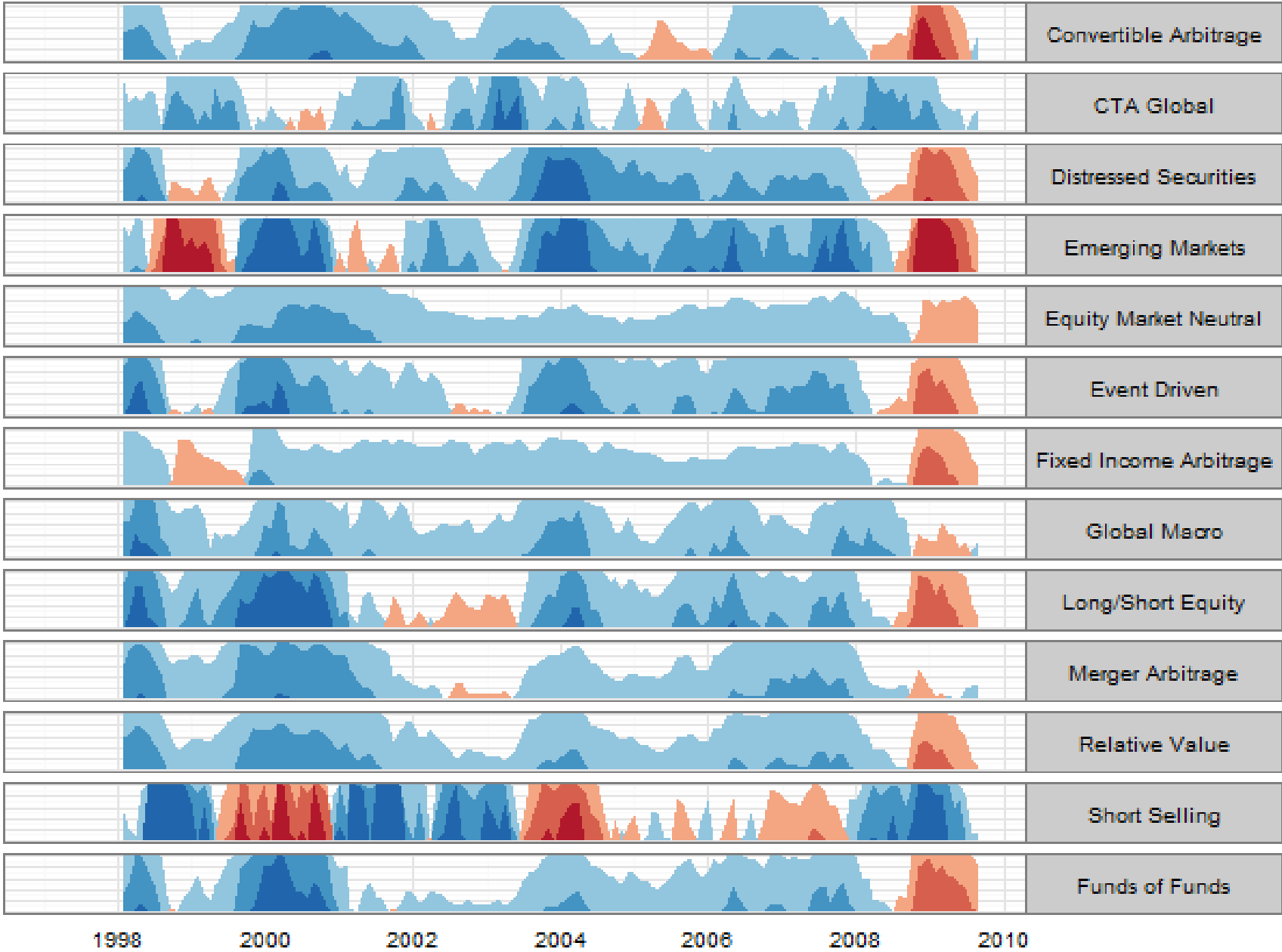
while

function

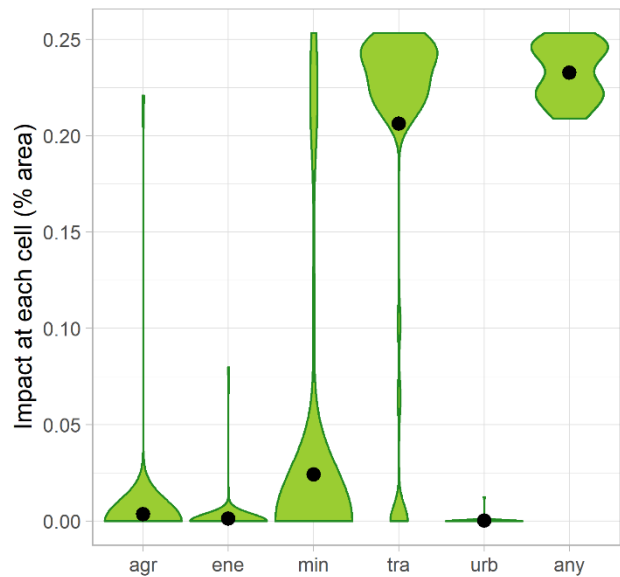




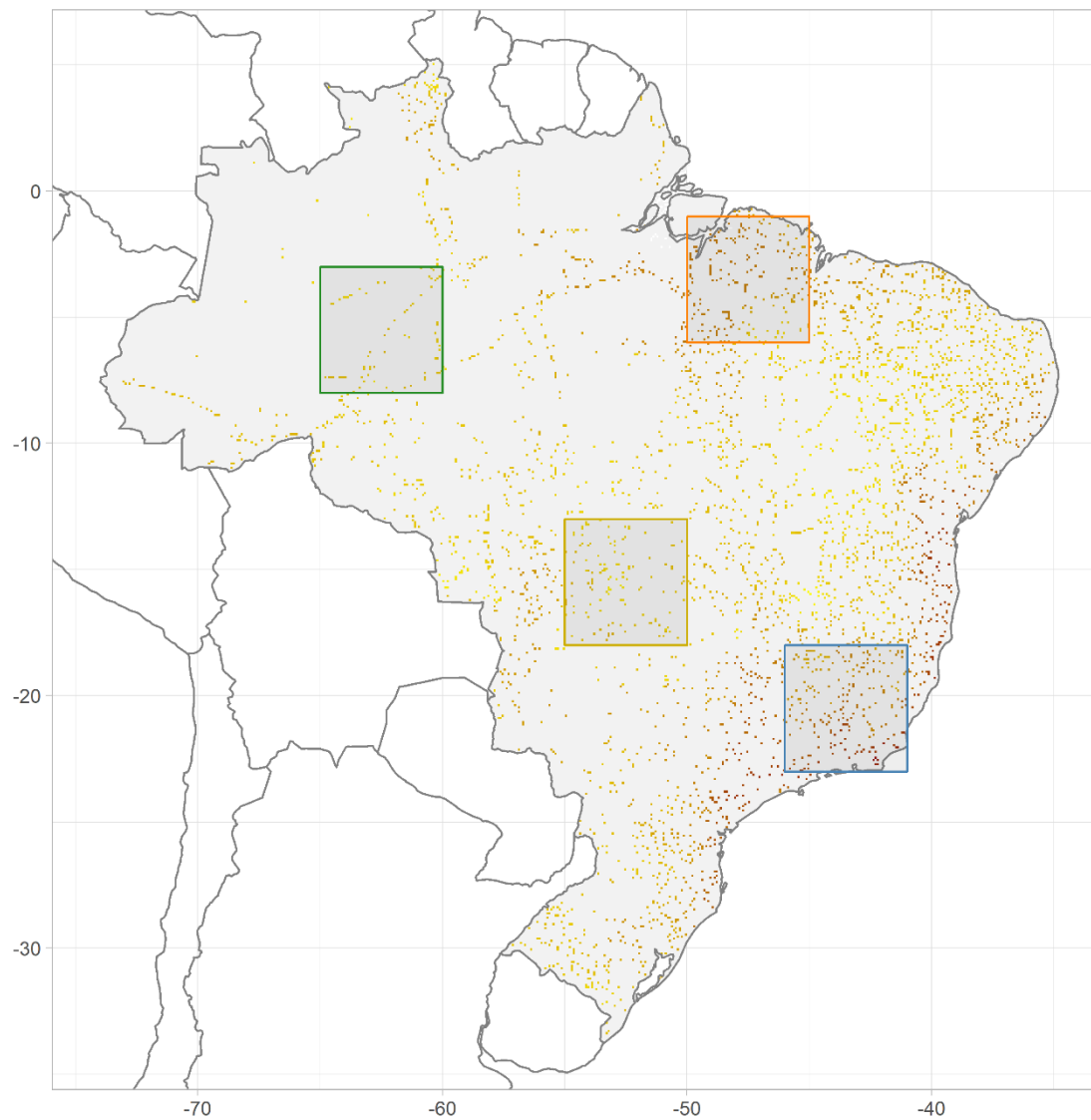
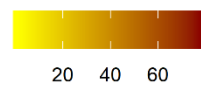
EDHEC Indexes Return (Rolling 1 Year)



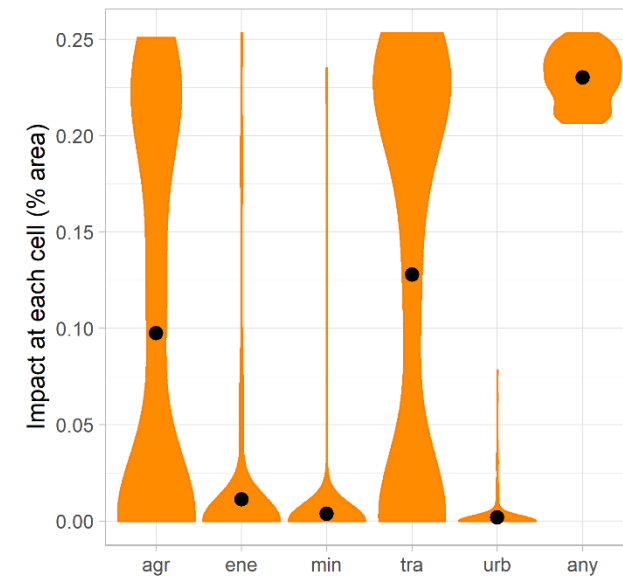
Amazon



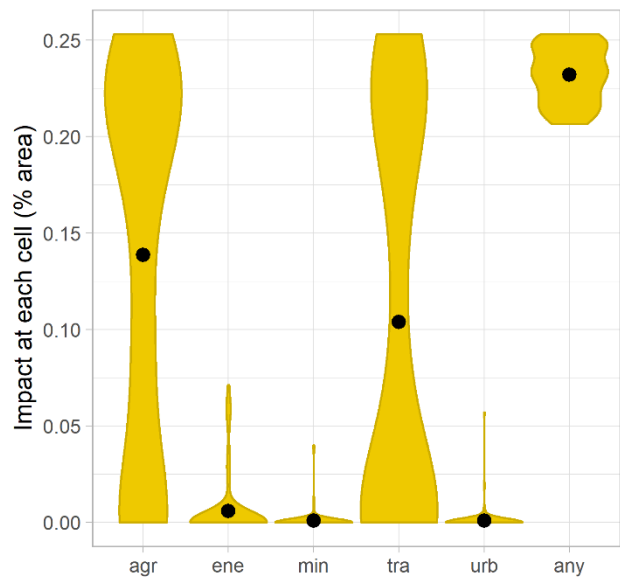
Threatened species richness



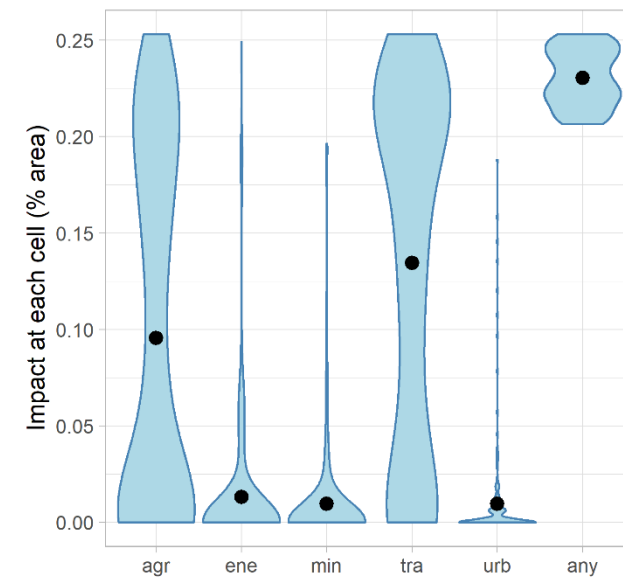
Northeast



Cerrado



Southeast



2- Data Visualization

- BLOCK 1 - INTRODUCTION TO GGPLOT
- BLOCK 2 – GEOMETRIES
 - BLOCK 2.1. - BASIC GRAPHS
 - BLOCK 2.2. - DISTRIBUTION GRAPHS
- BLOCK 3 - OPTIONAL LAYERS
- BLOCK 4 - SAVE FIGURES AND MORE
- BLOCK 5 - ANIMATED GRAPHS

BLOCK 1

Introduction to ggplot

BLOCK 1

Introduction to ggplot

1. Introduction to ggplot logic

History of ggplot2

- Developed by Hadley Wickham since 2005.
- One of the most-popular packages in R:
 - The second most-downloaded in the last month.
 - <https://www.r-pkg.org/downloaded>



Why is ggplot2 special?

- Versatility
- Logic
 - Formal, structured, clear and consistent
 - Based on **Leland Wilkinson's *Grammar of Graphics*** (that's why `ggplot`)
- Nice results!



But no one is always the best...

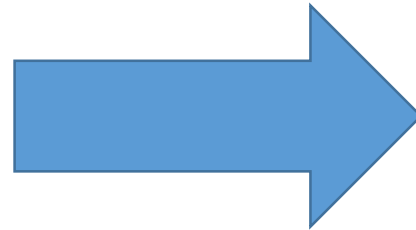
- R Base can be more adequate for a quick exploration, when data is not structured for ggplot2
- Some packages developed for specific graphics could perform better for these cases

...though... in most cases, ggplot2 offers the best results in relation to invested time, including figures ready for publication.

Data visualization

Data properties

- Numerical values
- Factor levels



Visual properties

- Point xy position
- Line colour
- Bar height
- ...

In Grammar of Graphics,
to represent a value by a colour or by point coordinates is conceptually equivalent.

Grammar of Graphics

Leland Wilkinson, Grammar of Graphics, 1999

2 CONCEPTS:

- *Grammatical elements* – “what” - graphical elements
 - Combined and structured in layers
 - Analogous to linguistic grammatical elements (subject, verb, ...)
- *Aesthetic mapping* – “how” - mapping rules
 - Define the relationship between data structure and graphical visualization
 - Analogous to linguistic grammatical rules

Data visualization

Data properties

- Numerical values
- Factor levels



Visual properties

- Point xy position
- Line colour
- Bar height
- ...

Grammatical elements
(graphical elements)

In *Grammar of Graphics*,
to represent a value by a colour or by point coordinates is conceptually equivalent.

Sintaxis in layers:

ggplot() + other functions

ggplot(**data**, **aes()**) +

← Basic ggplot function

geom_() +

← Other functions from ggplot2 package
(or from other packages associated)

facet_() +

theme_()

Element	Notation	Needed?	What it is?	Examples
Data	data	Essential	Database	Database
Aesthetics	aes()		Basic rules to transform data properties in visual properties	X-y coordinates Line colour Point shape
Geometries	geom_()		Geometrical objects that will represent data	Bars, lines, points
Facets	facet_()	Optional	Rules for multiple graphs	Number of lines and columns in the multiple graph
Statistics	stat_()		Data statistics to be represented	Sums, means, quantiles, adjusted curves,...
Scales	scale_()		Mapping rules in relation to scale	Colour scale
Coordinates	coord_()		Graph space	Coordinates, limits, ...
Themes	theme()		Graph appearance not related to data	General graph appearance, legend design, title design, ...

BLOCK 1

Introduction to ggplot

2. Building a simple plot in ggplot

```
ggplot(data, aes())
```

data

- in data.frame
- different data.frame can be used in the same plot:
 - One required in data argument of ggplot() function
 - Others can be specified in subsequent functions
- in “long format”

ggplot(data, aes())

aes()

- Define data visualization rules (“grammar”)
- aes() is a function required in the argument mapping of ggplot function:

```
ggplot(data, mapping=aes())
```

- aes() can also be specified as an argument of other functions:

```
geom_(aes())
```

If not specified in other functions, it is inherited from ggplot() function

geom_()

It defines the type of graphic.

Examples:

- geom_bar()
- geom_line()
- geom_area()
- geom_point()

Minimum code for a plot in ggplot:

```
ggplot(data, aes()) +  
  geom_()
```

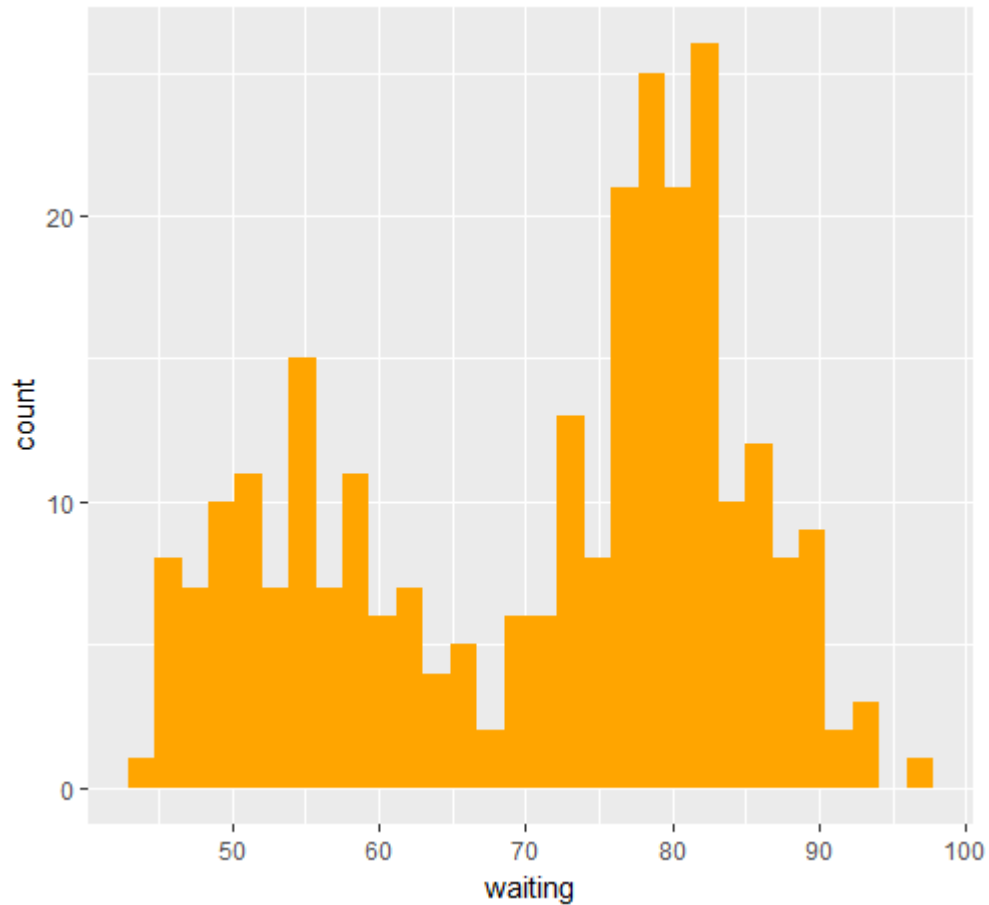
BLOCK 2

Geometries

BLOCK 2.2. - DISTRIBUTION GRAPHS

Observed frequency

Frequency of each value in a variable
(**OBSERVED** value)



Density function

Probability of each value in a variable
(**ESTIMATED** value)

