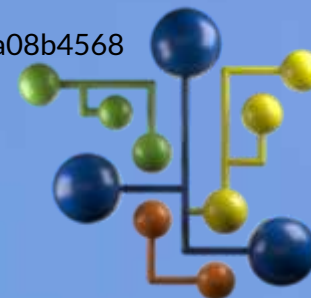


Data Science  
Academy

Data Science Academy gilbertogcs.santos@gmail.com 5c51e26c5e4cde78a08b4568



# Big Data Analytics com R e Microsoft Azure Machine Learning



Data Science  
Academy

Data Science Academy gilbertogcs.santos@gmail.com 5c51e26c5e4cde78a08b4568

# Big Data Analytics com R e Microsoft Azure Machine Learning

Manipulação de Dados com R

Seja Bem-Vindo(a)!



Data Science  
Academy

Data Science Academy gilbertogcs.santos@gmail.com 5c51e26c5e4cde78a08b4568

# Manipulação de Dados com R

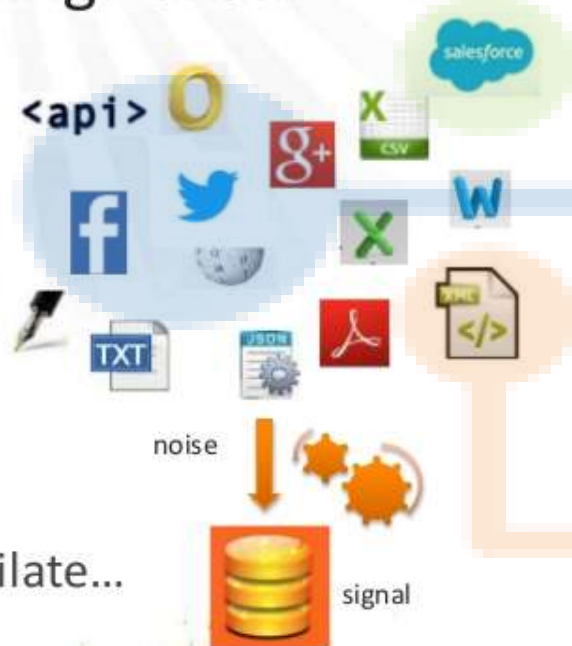
Data Wrangling



# Manipulação de Dados com R

## Data Wrangling: aka...

- Data Preprocessing
- Data Preparation
- Data Cleansing
- Data Scrubbing
- Data Munging
- Data Transformation
- Data Fold, Spindle, Mutilate...



A manipulação de dados pode ter muitas definições diferentes.

Mas o objetivo é um só: preparar os dados para o processo de análise, pois dificilmente os dados virão no formato ideal.





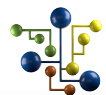
Data Science  
Academy

Data Science Academy gilbertogcs.santos@gmail.com 5c51e26c5e4cde78a08b4568

# Big Data Analytics com R e Microsoft Azure Machine Learning

O Que é Data Wrangling?

Seja Bem-Vindo(a)!



# O Que é Data Wrangling?

Preparação



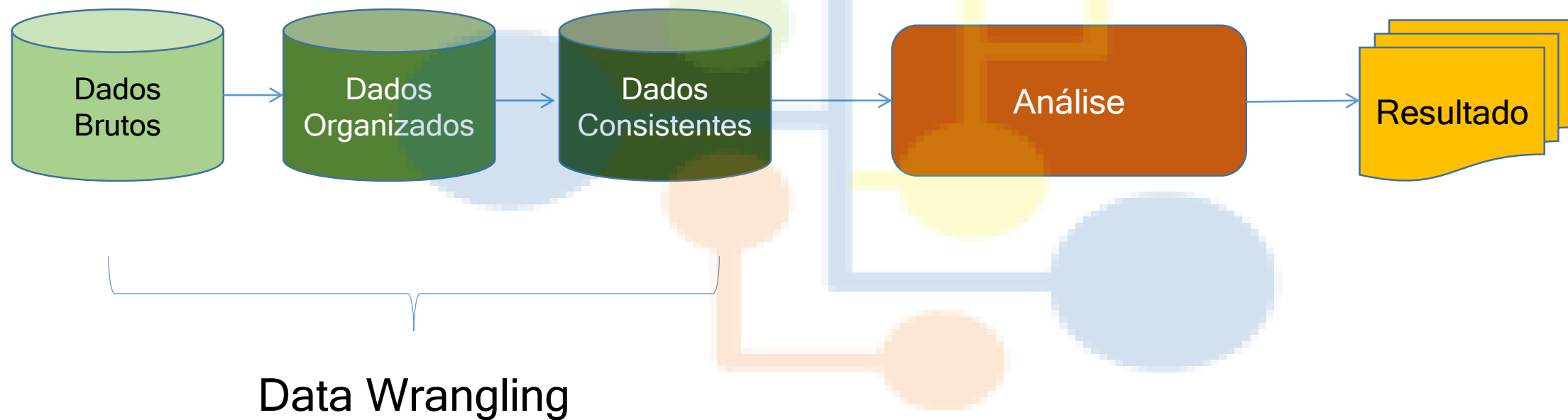


# O Que é Data Wrangling?

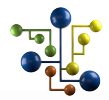
Data Wrangling  
(Manipulação de Dados)



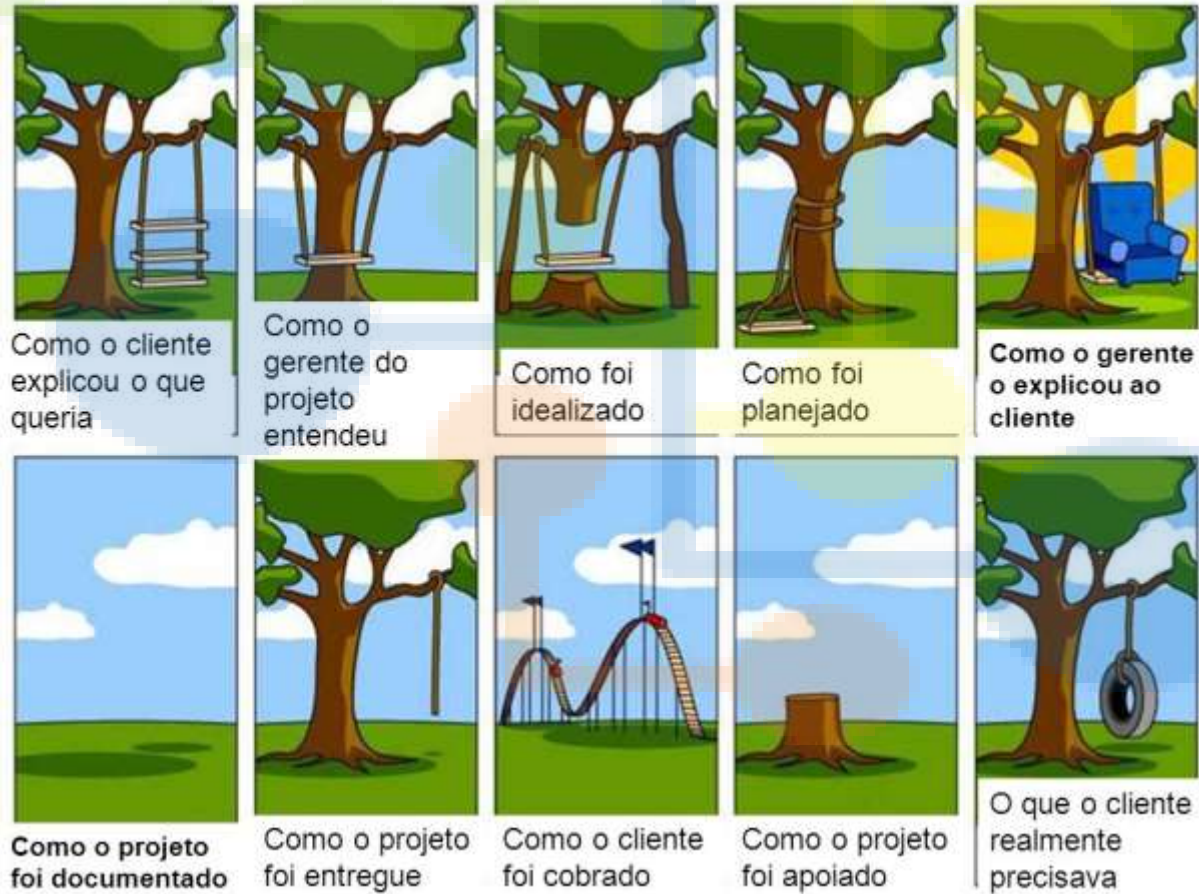
# O Que é Data Wrangling?







# O Que é Data Wrangling?





# O Que é Data Wrangling?

## Data Wrangling (Manipulação de Dados)

Limpeza, Processamento, Organização e Manipulação



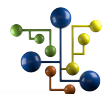
Data Science  
Academy

Data Science Academy gilbertogcs.santos@gmail.com 5c51e26c5e4cde78a08b4568

# Big Data Analytics com R e Microsoft Azure Machine Learning

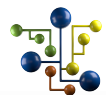
## Pacotes Para Manipulação de Dados em R

### Seja Bem-Vindo(a)!

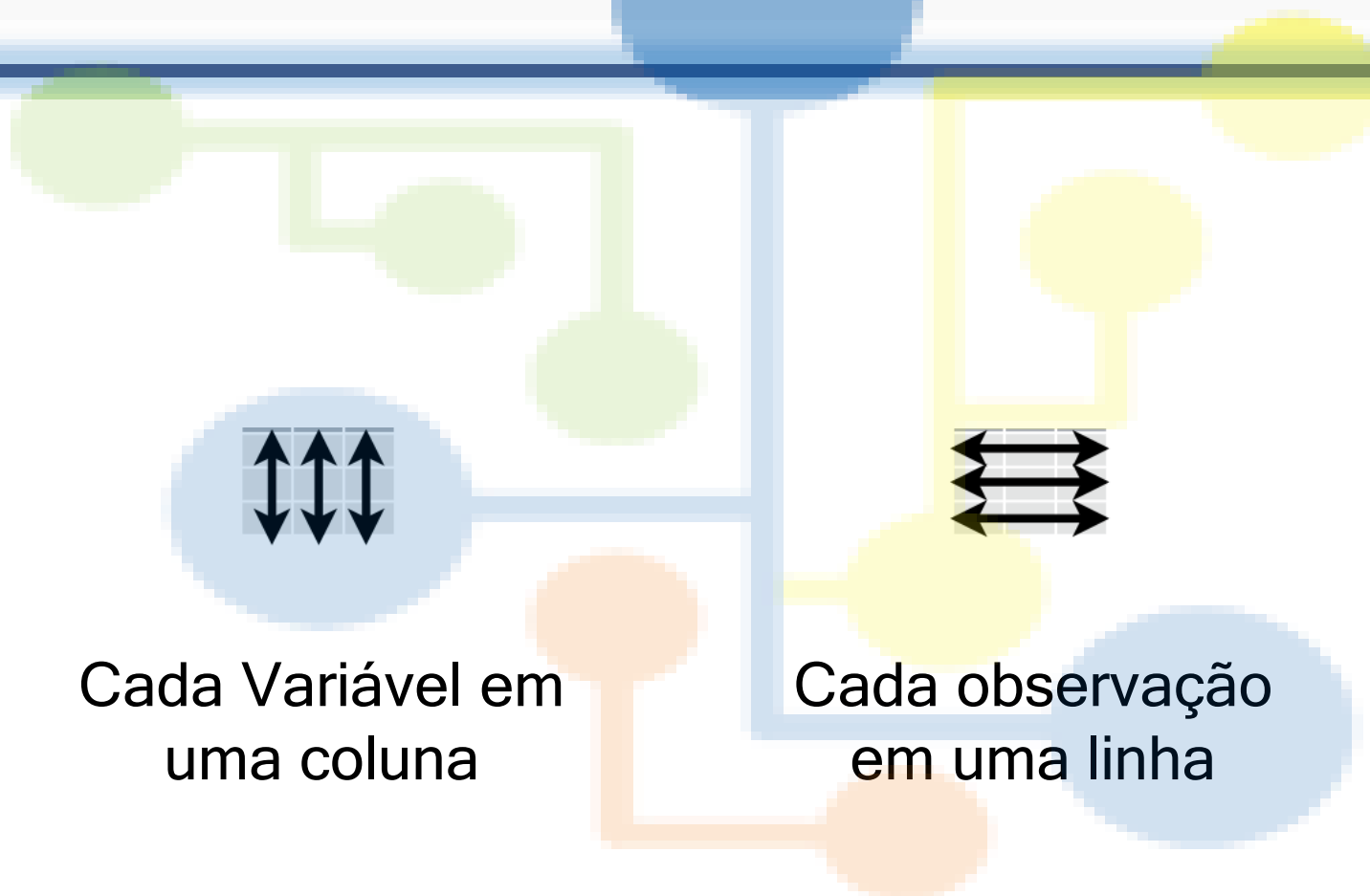


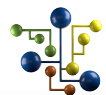
# Pacotes Para Manipulação de Dados em R

O que fazemos essencialmente durante a etapa de manipulação dos dados?



# Pacotes Para Manipulação de Dados em R





# Pacotes Para Manipulação de Dados em R

O que a Linguagem R pode fazer para  
ajudar o Cientista de Dados?



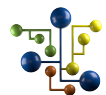
# Pacotes Para Manipulação de Dados em R

dplyr

- `select()`
- `filter()`
- `group_by()`
- `summarise()`
- `arrange()`
- `join()`
- `mutate()`

tidyr

- `gather()`
- `spread()`
- `separate()`
- `unite()`

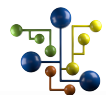


# Pacotes Para Manipulação de Dados em R

tidyr

Remodelagem de Dados





# Pacotes Para Manipulação de Dados em R

Funções tidyr

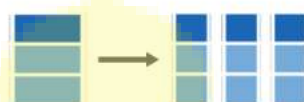


# Pacotes Para Manipulação de Dados em R

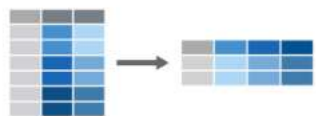
## Funções tidyr



`gather()`



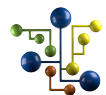
`separate()`



`spread()`



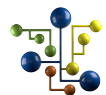
`unite()`



# Pacotes Para Manipulação de Dados em R

## Funções tidyr

Talvez você ainda não tenha percebido.  
Mas com apenas uma função, somos capazes de mudar completamente o formato (shape) dos nossos dados e isso pode fazer muita diferença no processo de análise.



# Pacotes Para Manipulação de Dados em R



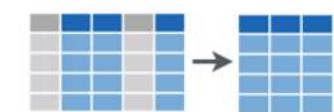
dplyr

Transformação de Dados

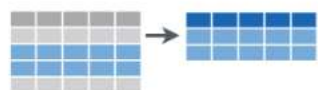


# Pacotes Para Manipulação de Dados em R

## Funções dplyr



`select()`



`filter()`



`group_by()`

summary  
function

`summarise()`



`arrange()`



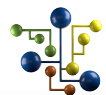
`mutate()`

x1	x2
A	1
B	2
C	3

+

x1	x3
A	T
B	F
D	T

`join()`



# Pacotes Para Manipulação de Dados em R



Por que vc não me mostrou isso  
antes?????



# Pacotes Para Manipulação de Dados em R

Porque antes de aprender a correr,  
é preciso aprender a andar.

E nossa jornada em análise de  
dados está apenas no começo!!!!





Data Science  
Academy

Data Science Academy gilbertogcs.santos@gmail.com 5c51e26c5e4cde78a08b4568

# Big Data Analytics com R e Microsoft Azure Machine Learning

Operador de Concatenação

Seja Bem-Vindo(a)!





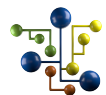
# Operador de Concatenação

Operador %>%

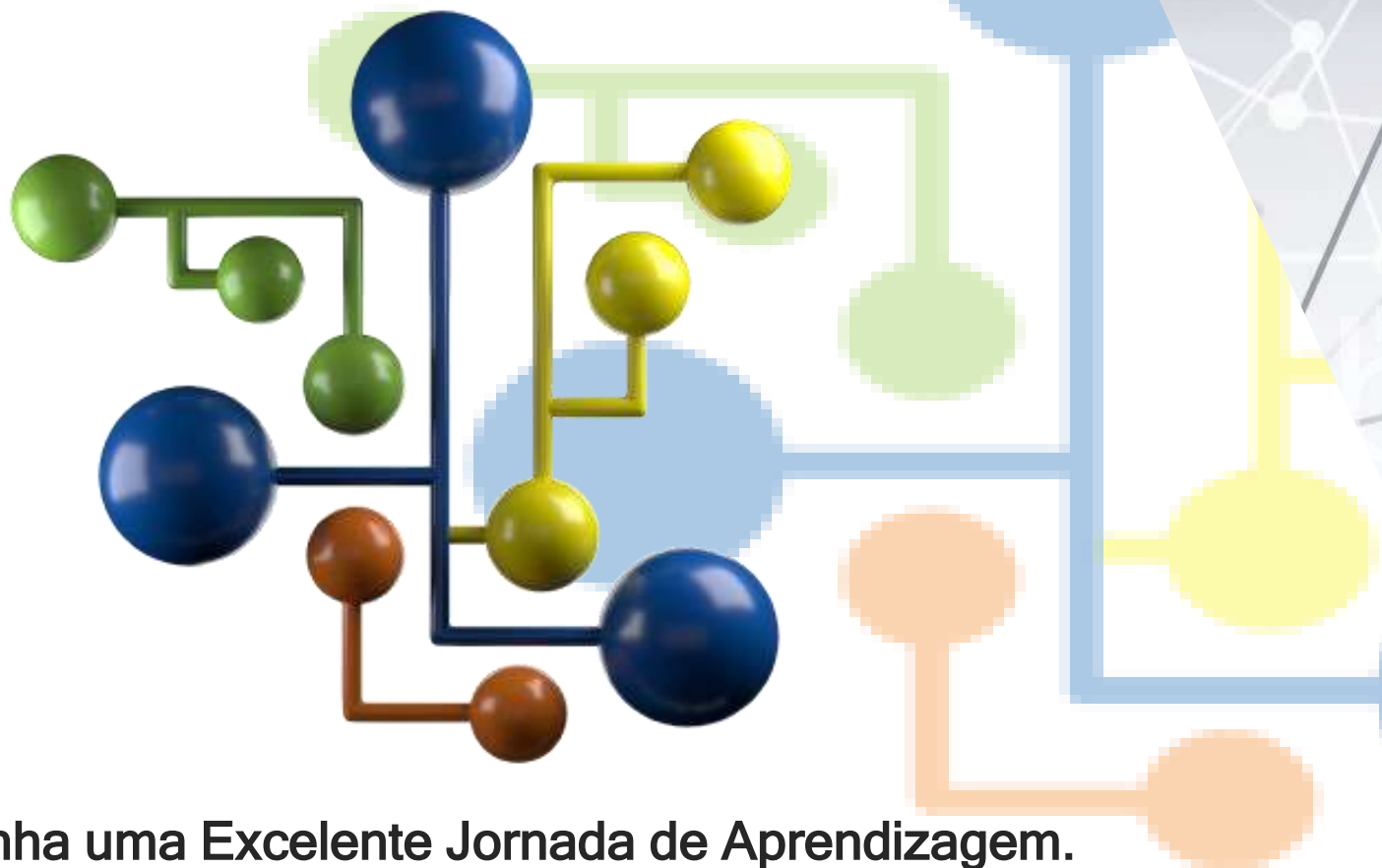
`filter(data, variable == numeric_value)`

ou

`data %>% filter(variable == numeric_value)`



# Muito Obrigado por Participar!



Tenha uma Excelente Jornada de Aprendizagem.

Equipe Data Science Academy