

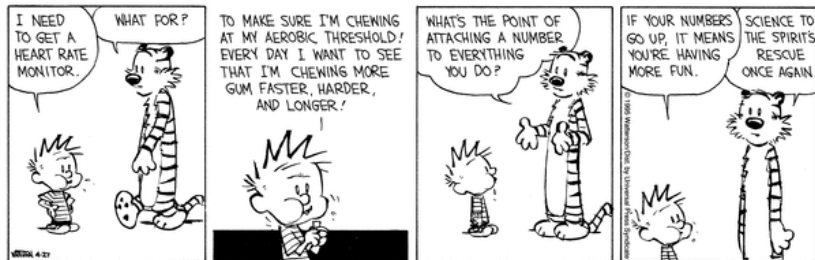
Data Science in R

Arthur Delarue

Overview of data science sequence

Today	Introduction	Arthur Delarue
Thursday	Data Wrangling and Viz	Xiaoyue Gong
Tue Jan 15	Machine Learning	Zach Blanks
Thu Jan 17	Advanced Topics	Phil Chodrow
Tue Jan 22	Project Presentations Deep Learning	Andreea Georgescu Zach Blanks

Data Science



Avoiding pitfalls



Genealogy of R

1. Scientists at Bell Labs invent 'S', the first statistical computing language. The name is chosen as an analogy to C. It is one of the first languages to have a graphics output, allowing scientists to print out plots of their data.
2. R is created by a gentleman named Robert Gentleman and his colleague Ross Ihaka in the 90s.
3. RStudio (the dominant IDE for R) released its first public version in 2011.

Why R?

1. It's free and open-source!

Why R?

1. It's free and open-source!
2. It's (fairly) universal

Why R?

1. It's free and open-source!
2. It's (fairly) universal
3. Large community means better support

Why R?

1. It's free and open-source!
2. It's (fairly) universal
3. Large community means better support
4. Extensive package library!

Let's dive in!