



R or Python ? Dilemma, dilemma...

Elsa Lozachmeur, Idorsia

Nicolas Dupuis, Sanofi



R?

- Released in 1995
- Popular, open-source language for statistical analysis
- Extended by packages almost at will.
- Rstudio
- R Shiny

Python?

- First released in 1991
- Fully functional, open-source, interpreted programming language for general purpose
- Famous for Deep Learning and Machine Learning
- Variety of Integrated Development Environments

Why the dilemma?

- High dependence to SAS, costly and proprietary
- Emergence of Open Source alternatives, trend towards decreasing this dependence
- New programmers in Data Science are likely to start with R or Python, not SAS

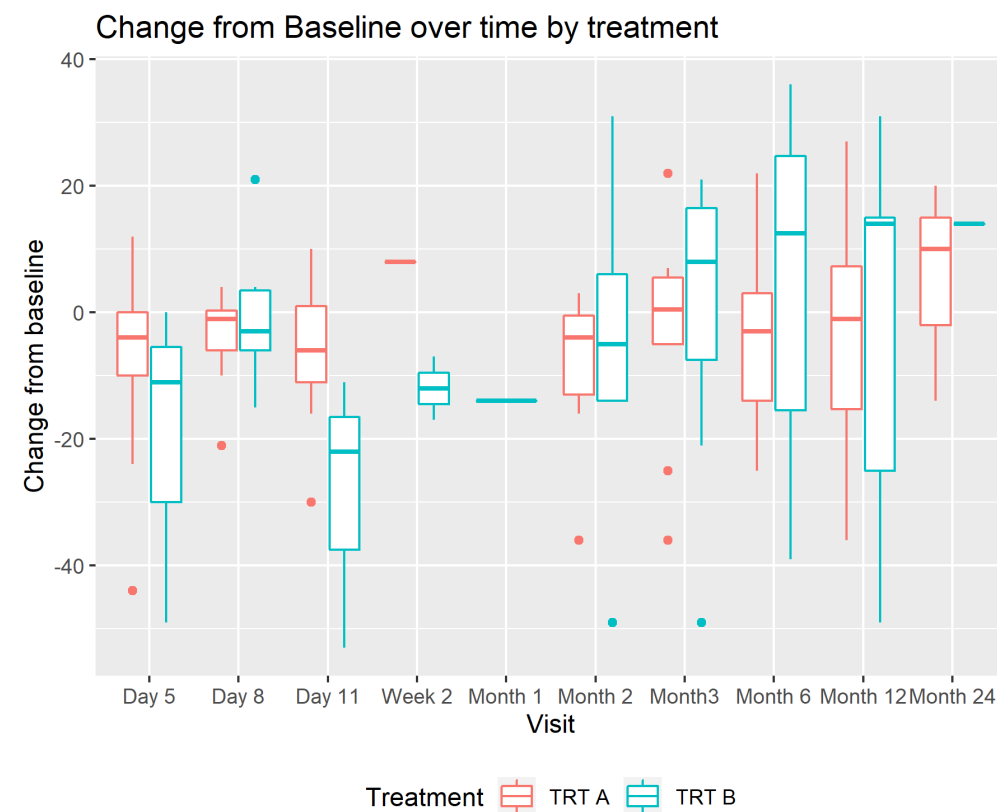
Our goal

We took a few classic SAS Statistical Programming tasks and tried replicate them in both R and Python

- explore data,
- derive some ADaM variables,
- create a table of summary statistics and a boxplot.

Our goal

		Treatment A			Treatment B		
Time point	Statistics	Baseline	Post-Baseline	Change	Baseline	Post-Baseline	Change
Baseline							
	n	xx			xx		
	Mean	xx.x			xx.x		
	SD	xx.x			xx.x		
	Q1	xx			xx		
	Median	xx			xx		
	Q3	xx			xx		
	Min	xx			xx		
	Max	xx			xx		
Month1							
	n	xx	xx	xx	xx	xx	xx
	Mean	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
	SD	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
	Q1	xx	xx	xx	xx	xx	xx
	Median	xx	xx	xx	xx	xx	xx
	Q3	xx	xx	xx	xx	xx	xx
	Min	xx	xx	xx	xx	xx	xx
	Max	xx	xx	xx	xx	xx	xx





Comparison

- Learning Curve
- Stability & Documentation
- IDE
- Performance
- Compliance

Conclusion

- Many pros and cons
- Current trend to switch to R for data manipulation and visualization
- Python more and more used for data science, machine learning and applications



What about
your
experience ?

