Programming and frameworks for ML (Python/R)

Daniel Villanueva
Benjamin Kraska

About Me

- Big Data Consultant at Indra / Big Data Lecturer
- More than 20 years of experience in different environments, technologies, customers, countries ...
- Passionate data and technology
- Enthusiastic Big Data world and NoSQL

Daniel Villanueva Jiménez
Consultor BigData en Indra
INDRA • Universidad Pontificia de Salamanca

▶ daniel.villanueva@immune.institute

Agenda

April 4 – Session 4	Daniel Villanueva	Introduction to R
April 17 – Session 5 April 18 – Session 6	Daniel Villanueva	Data Manipulation with Python
April 24 – Session 7 April 25 – Session 8	Daniel Villanueva	Introduction to NoSQL
April 25 – Session 8 May 8 – Session 9	Benjamin Kraska	Exploratory analysis

Agenda

- **Introduction to R**
  - Intoduction
  - Installing R
  - Data Structures
  - Statistical functions
  - Probability Distributions
  - Writting functions
  - Graphics in R

Agenda

- **Python: Data Manipulation with Pandas**
    - Numpy
    - Pandas
    - Dataframes
    - Reading / Writing Data
    - Exploring a DataFrame
    - Operations on a DataFrame

Agenda

- **Introduction to NoSQL Databases**
  - Case of use
  - Relational Databases
  - NoSQL
  - Riak
  - MongoDB
  - Apache Cassandra
  - Neo4J

Agenda

- **Introduction to exploratory analysis**
  - Descriptive statistics
  - Group by
  - ANOVA
  - Correlation
  - How to do it in Python / Pandas

THANKS FOR YOUR ATTENTION

Daniel Villanueva Jiménez
daniel.villanueva@immune.institute

@dvillaj