NATIONAL RESEARCH
UNIVERSITY

Big Data Myths and Facts:
Explaining Digital Transformation to non-IT Professionals

Boris Novikov
National Research University Higher School of Economics
Saint Petersburg, Russia

Myths are Everywhere

- Misterios millenium
- Software engineering myths
  - Performance is not an issue
- Myth is a misplaced, over-generalized, mis-interpreted, or mis-used fact

Digital Literacy and Digital Culture

- Top-Down initiative
    - The whole population considered digitally illiterate
    - Enforcement of digital economy
- All students of the St. Pegersburg university must take a course on digital culture

Saint Petersburg State University: Schools (Incomplete and imprecise)

- Law
- Journalism
- History
- Philosophy
- Arts

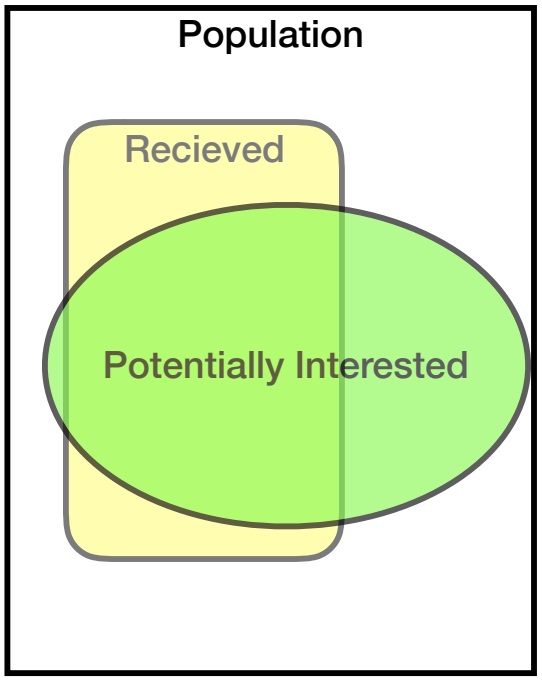
- Economy
- Management
- Liberal Arts

- Linguistics
- Social Sci
- Psychology
- Medicine
- Biology
- Geology

- Physics
- Chemistry
- Math & Mech
- Applied Math
- Math & CS

An Example: ADs Distribution

- A family complained on offending ADs
- The sender apologized and refered to an error in data analysis
- Few months later the claim was cancelled
- Mass media:
    1. Theny know more about us than we do
    2. Security must be improved
- Professional:
    3. Sometimes data analytics mey provide correct results
    4. Precision?
    5. Recall?



Responsible and Irresponsible Data Science

- SIGMOD 2019 Panel on responsible data science
- Examples of irresponsible data analytics
    - Face recognition
    - Identifying criminals
    - Gender recognition
- Failures of machine learning
- Interpretability

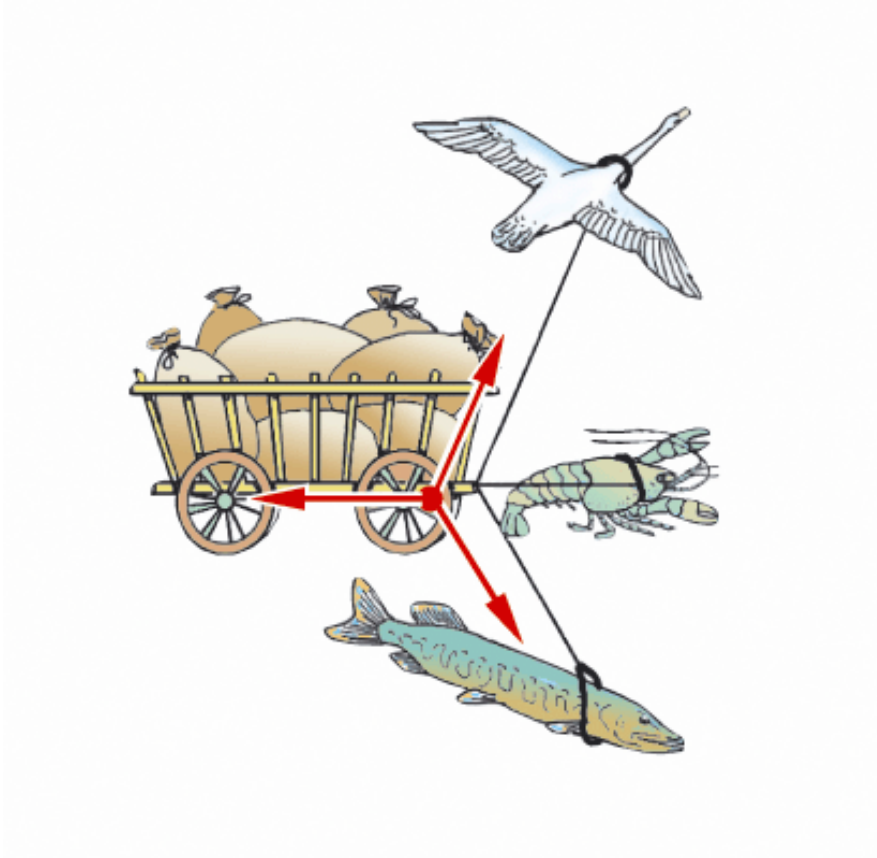
Digital Culture: Making Sense of it

Ideally, the course should address the following:

- How big are big data?
- Collecting data
- Analyzing data
- Evaluating the results
- When to involve data analytics professionals?

Developing the Course: the Team

- Creating a mandatory course for thousands of students
- Representatives of all schools
- Working group included 34 persons
- Diversity of opinions
- A set of slideshows with recorded lecutrer's voice



Course Topics

- The future is digital: official regulations, programs, declarations, etc.
- Internet resources
- Security and privacy
- Basics of statistics
- Data analytics, machine learning, and artificial intelligence
- Introduction to technologies

Presenting the Content

- Avoiding both complications and over-simplifications
- Popular presentation, but not a cookbook
- Avoid "knowledge for dummies" style
  - Avoiding "Easy, do it yourself"
  - Positive template: "Basic models of nuclear physics may be presented, but do not try to explain how to make nuclear weapons in your kitchen"

Conclusions

- Myths are widespread
- Probably it is already too late
- We still have to try to educate