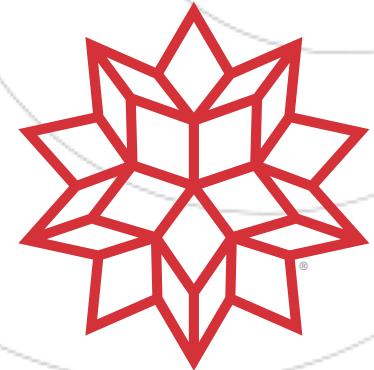


Multiparadigm Data Science

Erez Kaminski
Technology Specialist
Wolfram Research
Twitter: @Erez_Kaminski



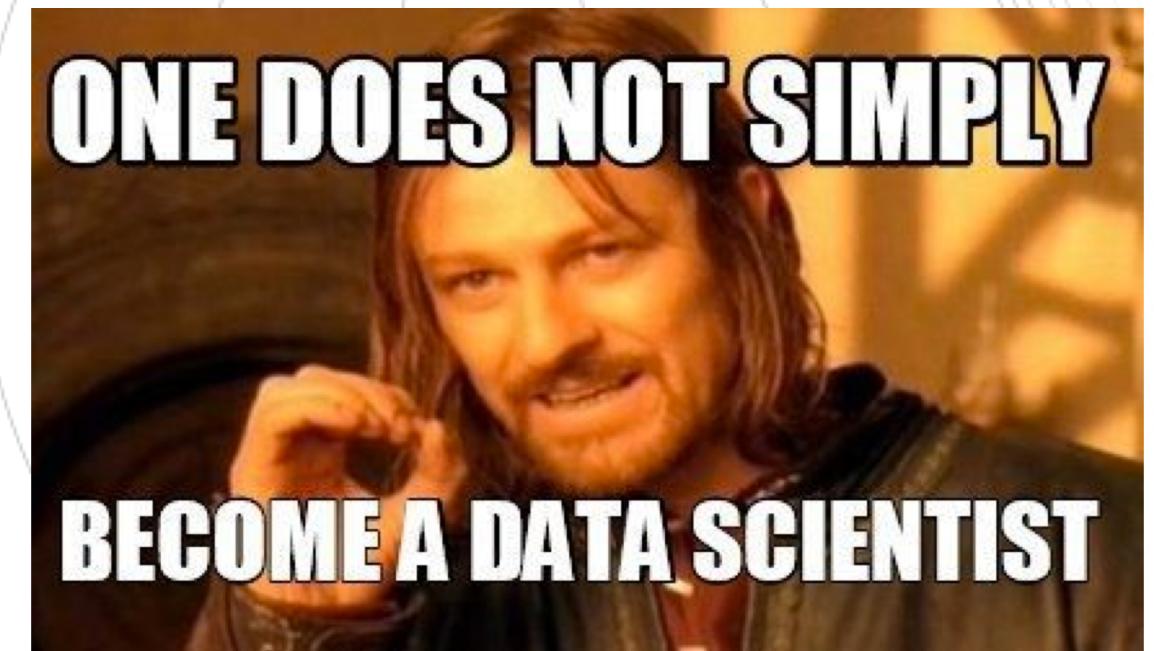
WOLFRAM
COMPUTATION MEETS KNOWLEDGE

Talk Outline

- **What is Data Science?**
- **Multiparadigm Data Science**
- **Example of a Data Science Workflow**
- **The Data Science Pipeline (full demo)**

What will you get from this workshop?

- Learn about Multiparadigm Data Science
- Learn how to build a Data Science pipeline
- Learn about applying:
 - Machine Learning
 - Statistics
 - Importing and scraping
 - Visualization
 - Much more!



What is data science?

Data science is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms

"Data Science." Wikipedia, Wikimedia Foundation, 16 May 2018, en.wikipedia.org/wiki/Data_science.



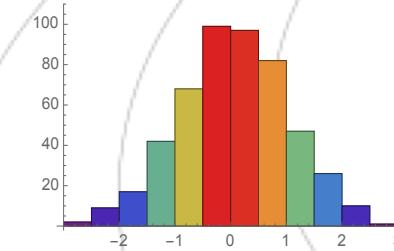
Current Challenges in Data Science

- Handling different types of data
- Slow insight cycle and reproducibility
- Lack of communication and feedback from stakeholders
- Inability to interface with results
- No organization-wide deployment

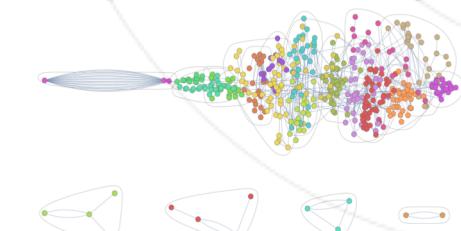
What is Multiparadigm Data Science?

- Data Science that:
 - Uses flexible pipelines built on arbitrary input
 - Delivers better and faster insight
 - Uses meta-algorithms
 - Is highly automated
 - Scalable
 - Is infused with real-world data
 - Delivers the right output

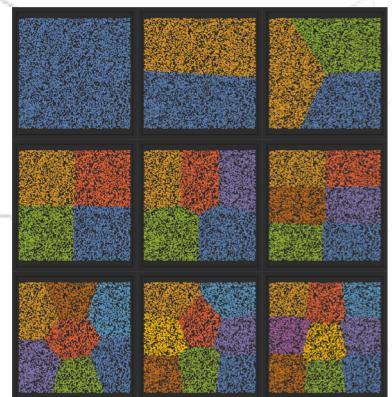
Data Visualization



Graph Analysis

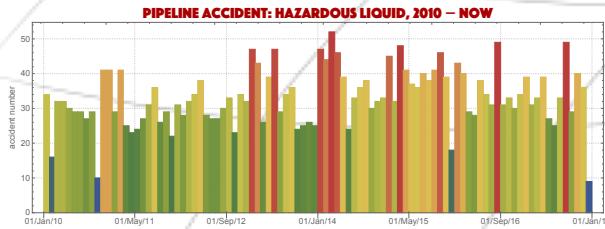


Data Clustering

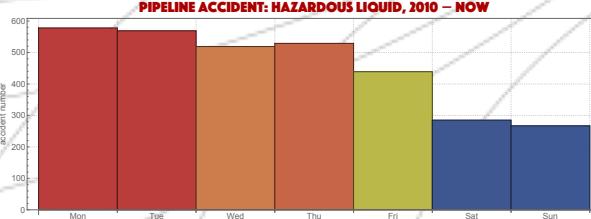


A Quick Example

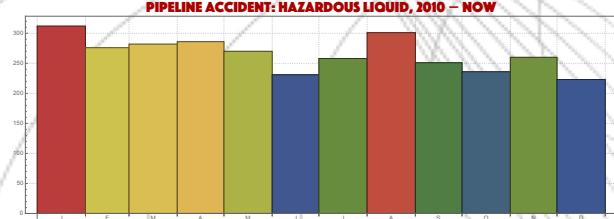
Year and Month



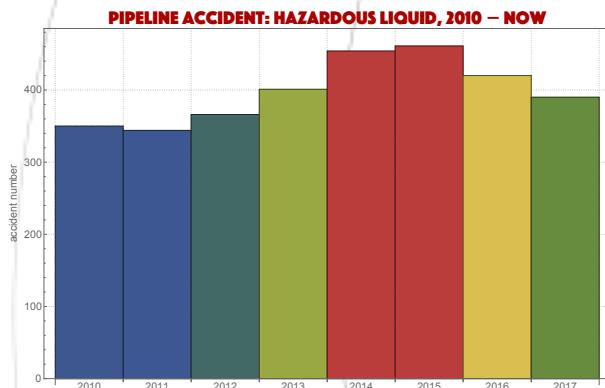
Weekday



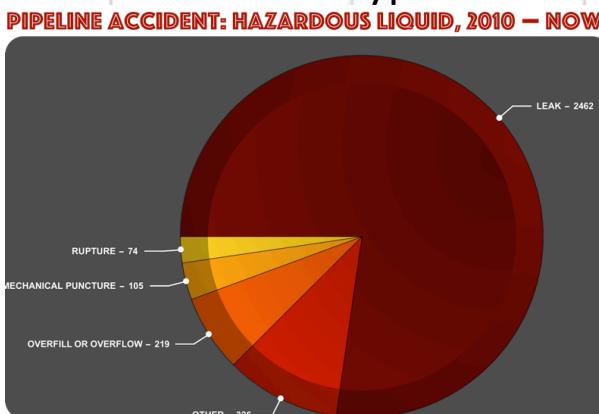
Month



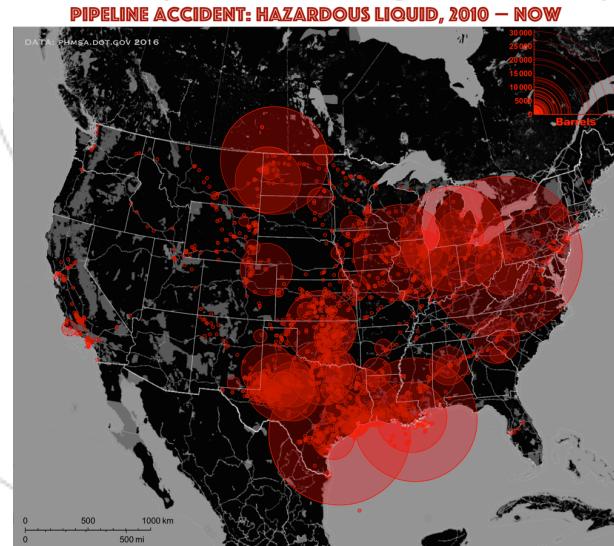
Yearly



Incident Type



Location



Basic Workflow

Step 1

- Get the Data

<https://data.cityofchicago.org>

Step 2

- Process and Explore Data (standardize)

Notebook

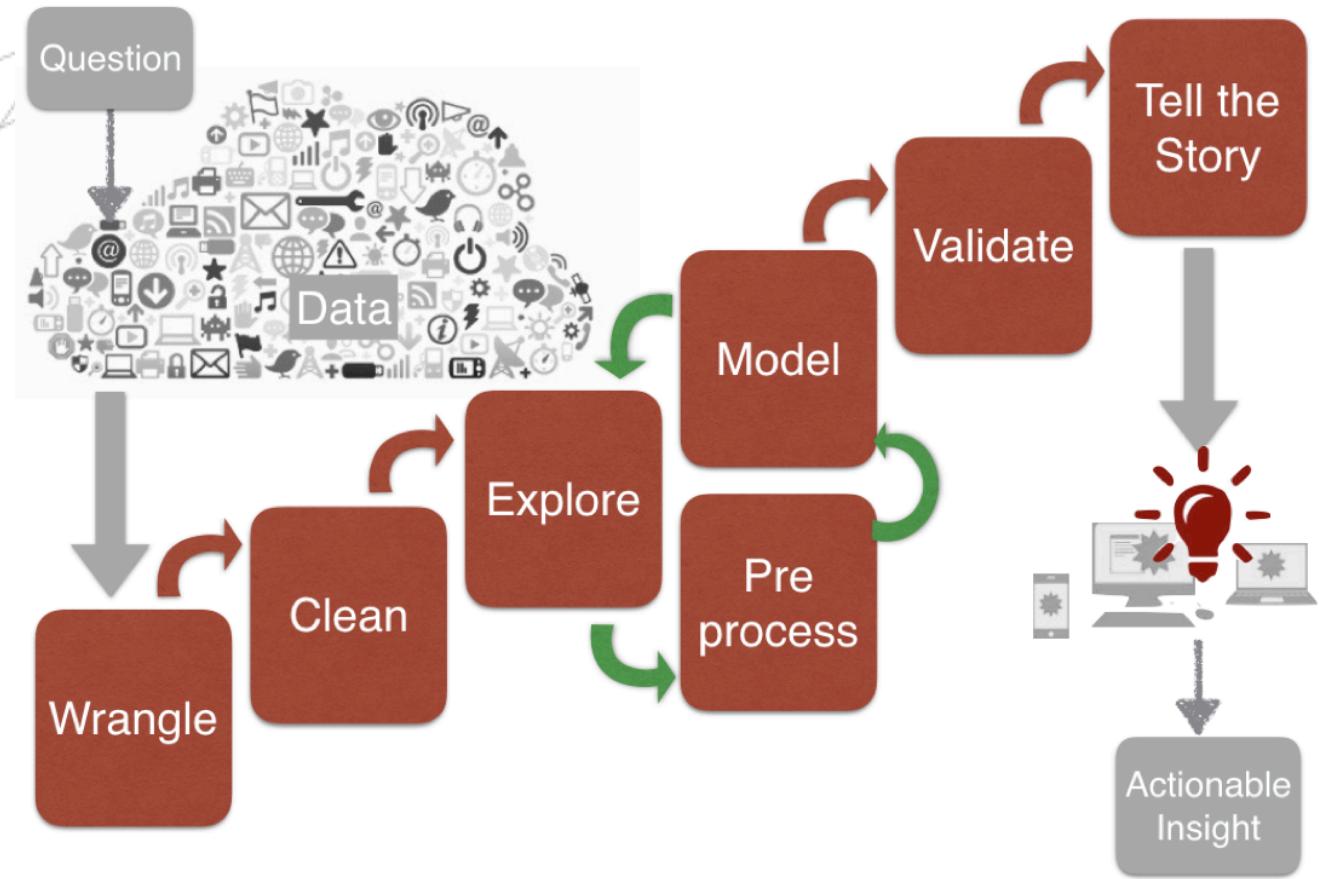
Step 3

- Deploy the Data

The Data Science Pipeline

- Complete Adaptable Workflow
- Systematic
- Agile
- Reproducible
- Standardized output

[Demo](#)



Conclusion

- Multiparadigm Data Science allows you to:
 - Handel different types of data
 - Explore & explain the data in different ways
 - Deploy it everywhere
 - Create a long-term pipeline

www.mpdatascience.com

Erez Kaminski

Twitter: @erez_kaminski

Email: erezkaminski@wolfram.com