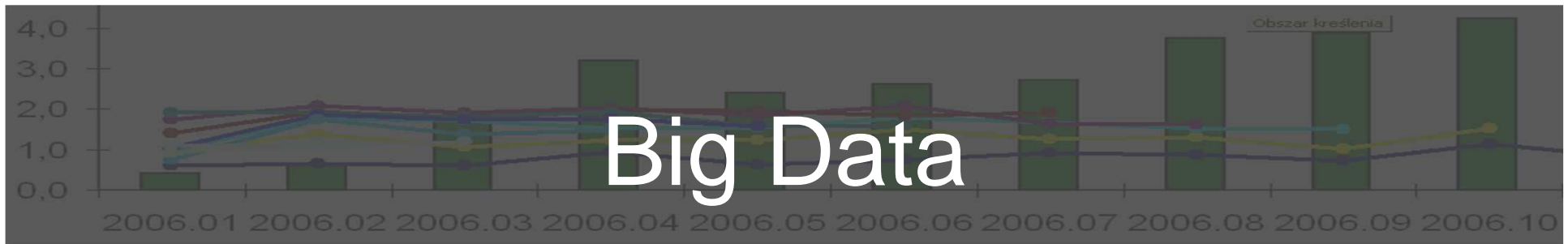


Warsaw School of Economics – SGH
Institute of Statistics and Demography
Event History Analysis and Multilevel

Introduction to data processing

Dr. Karol Przanowski



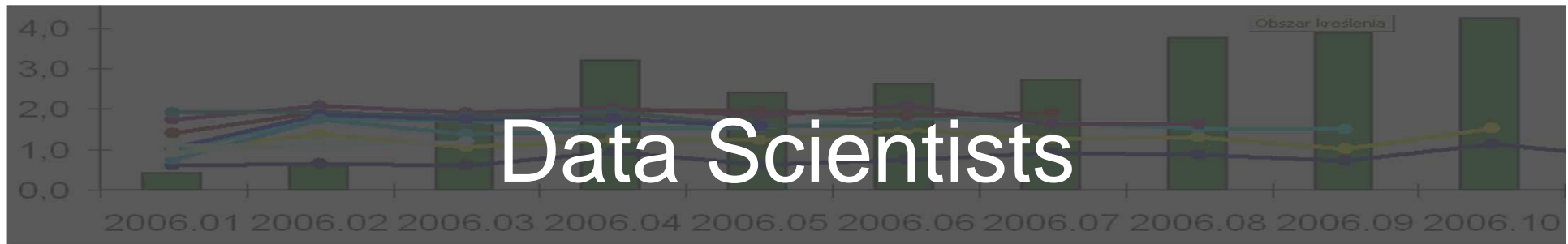
- Was born in 2001
- 3V then 5V:
 - **Volume** – big data, massive,
 - **Velocity** – many changes in the time,
 - **Variety** – different structure, also unstructured,
 - Veracity – telling the true
 - Value – having information, useful.
- **In general to build predictive models**
- **Process optimization (on-line)**

<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>



- General message:
 - Let us to speak data, listen to data
 - We do not need to know, if it works
 - (not why, but what)
 - If we can earn money, all is ok!!!
 - Let us gather the data even we do not know its usage

Mayer-Schonberger V i Cukier K (2013). *Big Data: A revolution that will transform how we live, work and think*. An Eamon Dolan Book, Houghton Mifflin Harcourt, Boston, New York.

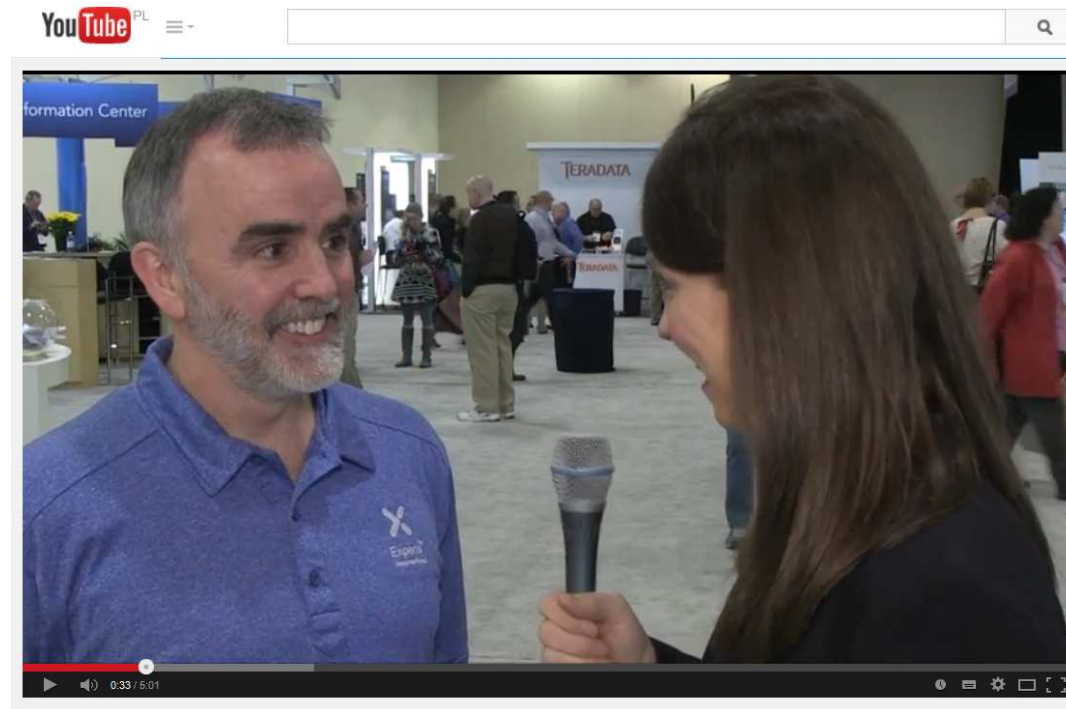


- Data Scientist – the most sexy job in XXI (Harvard Business Review 2012)
- Many skills in one person:
 - Statistics, IT, Business
 - Communication, creativity
- Data transformation into important knowledge
- Data is the most valuable assets

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

http://manager.inwestycje.pl/manager_360/Rafal-Wojdan;244301;0.html

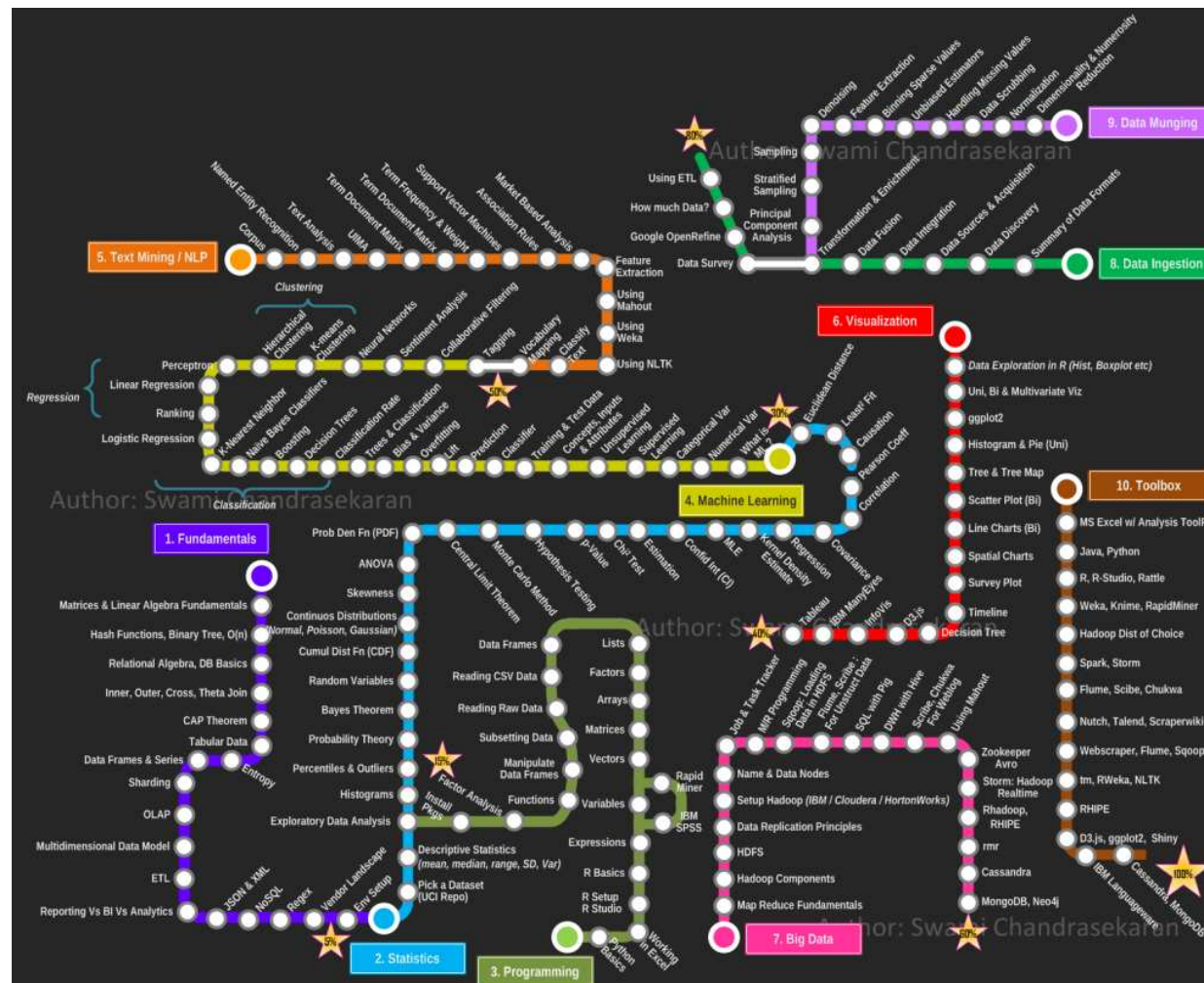
How to be a Data Scientist Using SAS®



Chuck Kincaid of Experis Business Analytics

<http://support.sas.com/resources/papers/proceedings14/1486-2014.pdf>

<http://www.youtube.com/watch?v=KLO9f7nx3Yw>



<http://nirvacana.com/thoughts/becoming-a-data-scientist/>

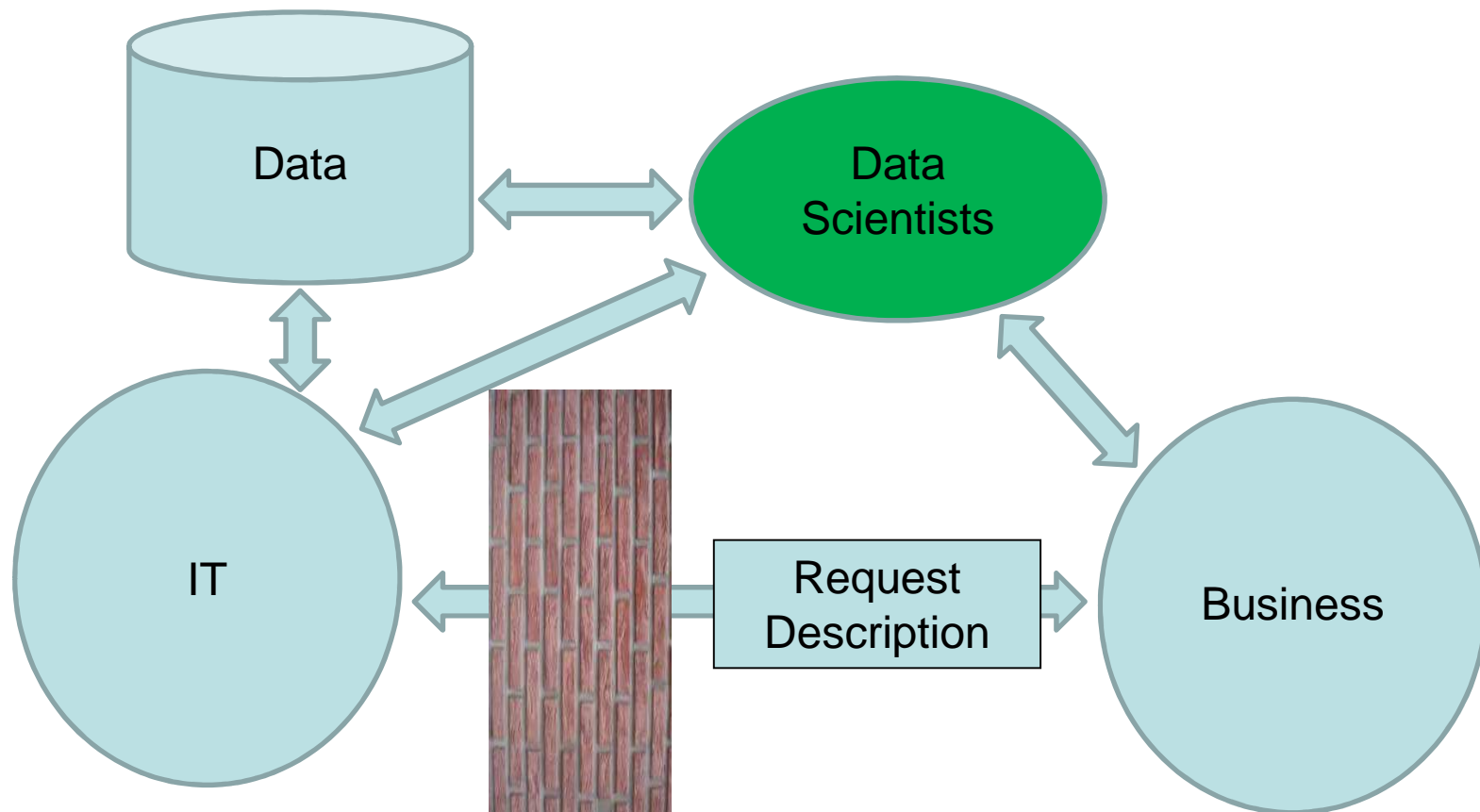


Data Scientists - competences

- Data processing (programming):
 - C++, Java, Python, Perl, R, SAS 4GL
- Systems:
 - Oracle, Teradata, SAS, Hadoop
- Statistics and Data Mining:
 - Logistic regression, tree decisions, neural networks, random forests, cluster analysis, survival analysis, CLTV models
- Text Mining

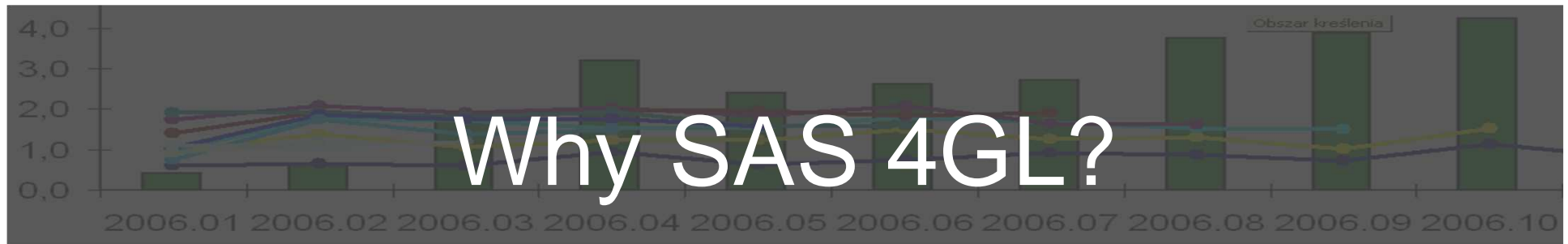
The role of Data Scientists

- Middleman, connector, between IT and business





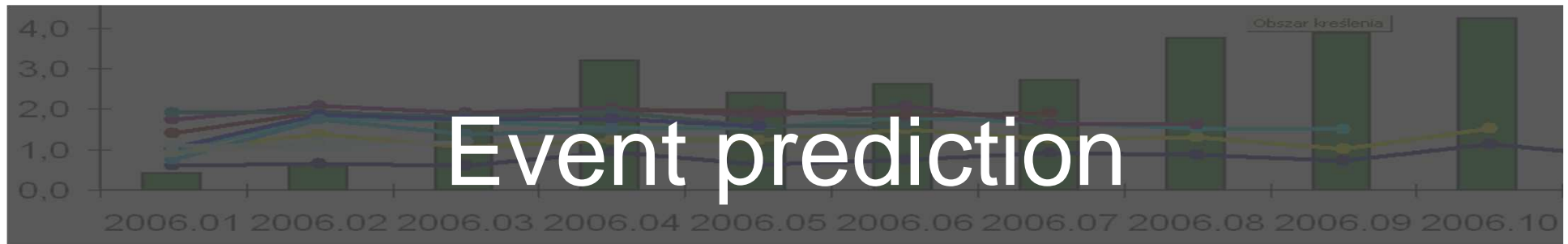
- Fully integrated system
- Exists in various companies: banks, insurances, Telco's
- Supports sciences: data science, medicine, biostatistics, economy etc.
- Helpdesk on high level and quality
- **S**tatistical **A**nalysis **S**ystem



- Elastic, easy in writing (typing, coding)
- Access into all data structures, formats
- Can be installed in various operating systems
- Microprogramming has a power
- Easy to find sample SAS programs on web sides
- Many conferences and blogs, including SUGI



- Repeatable and massive events
- Trend and property indication, discovery
- Population research
- Relation analysis
- Forecasting and predictive analysis
- Stability testing
- Not one event but a few thousands

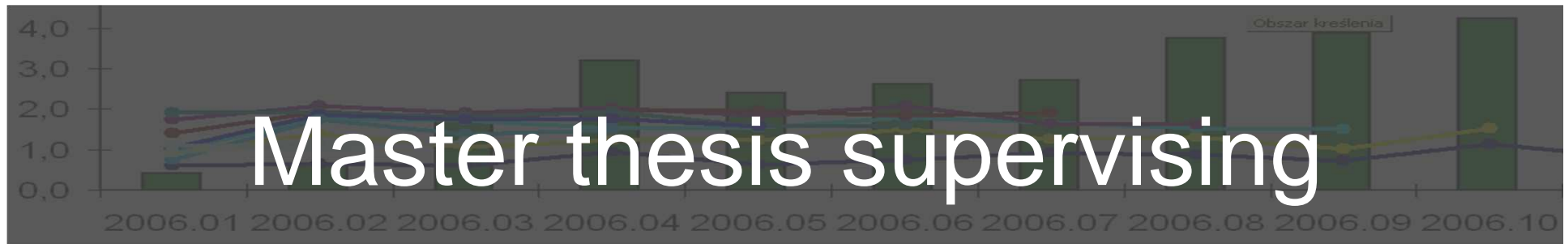


- New purchase
- Conversion into new product
- Instalment or credit payment
- Attrition, Churn
- Fraud, cheater, scam (AML)
- Not legal usage of electric service
- Accident, emergency event



Event History Analysis and Multilevel

- 2013 (International Year of Statistics 2013 www.statistics2013.org)
 - Advanced Analytics and Data Science
www.analytics-conference.pl
- 2014
 - II Advanced Analytics and Data Science – 14.10
http://www.sas.com/pl_pl/events/2014/advanced-analytics-and-data-science/index.html
- 2015
 - III Advanced Analytics and Data Science – 20.10
http://www.sas.com/pl_pl/events/2015/advanced-analytics-and-data-science/speakers-and-panelists-2015.html



- Scoring techniques and methods comparison
- Variable coding, binning
- Collinearity
- Reject Inference, MKS and MIV
- Crisis prediction and analysis, survival analysis
- Relation between predictive power and financial profit
- Model stability in the time
- Pricing management
- Variable monotonic property analysis