Yxilon - The Next Generation of Statistical Software

Wolfgang Härdle
Sigbert Klinke
Uwe Ziegenhagen

Institut für Statistik and Ökonometrie
Humboldt-Universität zu Berlin
http://www.quantlet.org
http://www.xplore-stat.de

Contents of This Talk

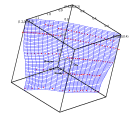
Introduction ✓

Statistical Software Packages

Requirements for Statistical Software Applications

"Each new generation of computers offers us new possiblities, at a time when we are far from using most of the possibilities offered by those already obsolete."

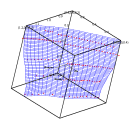
John W. Tukey (1965)



Statistical Software Packages

ACE, Alpha+, Analyse-it for Microsoft Excel, AnswerTree 3.0, Ascribe, 2ask, Askia, auditHOST,[B/D], Bellview® CAPI, Bellview® CATI, Bellview® Fusion, Bellview® Scan, Bellview® Web, Beyond 20/20, Beyond Question, Blaise, BMDP, Brand2hand,Clementine, CleverForm, Clicktools, Collect Data Now!, Community Express, Complete Report Automation, Confirmit, Cost$Benefit Analysis Tool, CSPro,DataSet V, DatStat Illume, Demographix, Dub Interviewer,Easyresearch - online surveys, Entryware™ Pro, EquivTest, ESBPDF Analysis, ESBStats, E-Tabs Enterprise, E-Tabs Interactive, E-Tabs Reader - Professional / Lite, E-Tabs Writer,Fitter, ForexastX Wizard, Formic 3, FUSE Enhanced Survey Solutions,GenStat, gllamm, Grader, GS+ Geostatistics for the Environmental Sciences,Halogen eSurveyor, HostedSurvey,IDAMS, IdeaMap.NET, Inquisite, InSite, Interviewer CAPI, Interviewer CATI, Interviewer VCC, Interviewer Web, Ioxphere, IRQuest, ISSA, IT, IYCQ Manager, IYSS,Jambo,KeyPlan, Keypoint 4.0, KnowledgeWorx,MaceTech Options Analysis, MacTech Blockage Analysis, Manet, MARSC, Mentor, Merlin, MERLIN FASTAB, MERLINPLUS, MERLIN TOOLKIT, Microtab XP, MIM, Mobile Researcher, MR Assistant, MRDCL, mrInterview, mrPaper, mrScan, mrStream, mrTables, mrTranslate, mSITE, mTAB,NCSS, Nesstar, NetTeam.Survey, NIPO Interview System, nQuery Advisor, NSDstat,Online Surveys, OnTraq, Opinio, Opinion Taker, OPST, Origin Survey Manager,PASS, Patient Profiles, Patient Satisfaction Survey Database, PerceptionStream, powerTABLE, Print & Scan, PRO-QUEST, Prospex, P-Stat, Pulsar, Pulsar Web,QDS, QEDML Designer, QEDML Web Server, QPSMR, QSR N6, QSR NVivo, Qualrus, Quanvert Publisher, QuestionPro, Questor,Raosoft EZSurvey, Raosoft EZSurvey SMS, REDATAM, Remark Office OMR, Remark Web Survey, Report Direct, Resampling stats, Research Reporter, Rogator G4, Rostock Survey Tool,SAS/STAT, SC, SENATE-Survey, SensorPro, SIR, snap Campus Edition, SNAP Internet module, snap Keystation Edition, SNAP Network Interviewing module, snap PDA Interviewer Module, snap Professional Edition, snap ProNet Edition, SNAP Results Edition, snap Scanning Module, snap SurveyPaks, SOLAS, S-Plus, SPSS 11.0, SPSS MR Data Model, SPSS MR Excel Viewer, SPSS WebApp, STAR®, Stata, Statistica, STATISTICA Data Miner, STATISTICA QC Miner, Statistix, StatPac for Windows, StatXP, StudyResult, StudySize, SumQuest, SuperSTAR, Survent, Survey Budgeting System, Survey Explorer, Survey Galaxy, SurveyGold Standard Edition, Surveyguardian, Surveyor, SurveyPro 3.0, Survey Quick, SurveySolutions, SurveyView, SurveyWriter, SYSTAT, SySurvey,Tables Direct, Team360 for Students, TeleForm, TextGrab, TEXTPACK, TextQuest, TouchForm, TPL Tables, TSP 4.5, TSS - The Survey System,Unistat,Vector, VentureFeedback.com, Verbastat, Visual QSL, Vitalnet,WARP-IT, WebMine, WebSTATISTICA, WebSurveyor, WebSurveyor 3.5, WinCATI, Winsteps, Winyaps,XLStatistics, XploRe

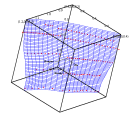
Figure 1: Association for Survey Computing Software Register



Microsoft Excel

Not listed in ASC directory but used for most analysis tasks

+ huge variety of import & export filters
+ wellknown interface, dynamic graphics
+ numerous extensible analysis functions
− tables limited to 65 536 rows $\times$ 256 columns
− numerical inaccuracies
− no explicit statistical displays



Excel

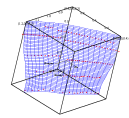
Excel is

- no dedicated statistics package
- used for majority of statistical analysis

What makes Excel so attractive?

- Organisation
- Analysis
- Presentation

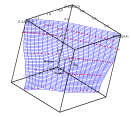
Apparently Excel performs well in at least one of these areas...



Requirements for Statistical Software

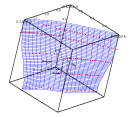
Chamber & Lang (1999):

1. Usable from multiple front-ends: Excel, webbrowser, DLL
2. UIs for different audiences: Each discipline has own culture of naming statistical phenomena, UIs must reflect this
3. Extensibility on language/interpreter and native level, e.g. conversion of highlevel to highspeed production code



Requirements for Statistical Software

4. Internet abilities to read and write data in networks
5. Database support to allow nearly real-time analysis
6. Interactive graphics and embedded controls to study the effects of parameter changes
7. Extensibility by inclusion of existing code
8. Optimization for performance (compiler level, multi-processor)

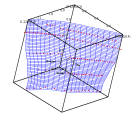


Further Requirements

Set of methods: one of the most crucial points for the commitment to a statistical software; common subset of methods but huge differences in configuration

Multiple language support: English is the lingua franca in science, but usability increases when userinteraction in native tongue.

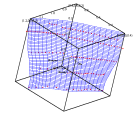
Valuable user ressources: printed/electronic manuals, tutorials and on-line



The Need for Manuals, Tutorials, etc. . .

"Most uses of the classical tools of statistics have been, are, and will be, made by those who know not what they do."

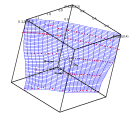
John W. Tukey (1965)



Satisfying the Requirements

Different software design approaches

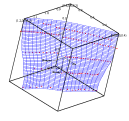
- one big monolithic application
- set of modules, each designed for special purpose



XploRe

- developed in a joint venture of Humboldt-Universität zu Berlin and MD*Tech
- C-style syntax, procedural approach
- available as Batch, standalone and Client/Server on Win32/Unix/Linux
- strong focus on non-parametric and quantitative finance methodology

Why do we need Yxilon?



Drawbacks of Monolithic Design

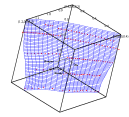
- implementation of new technologies increased complexity
- original design did not include e.g. transmission of data via internet
- all parts closely tied with each other

$\Rightarrow$ Reimplementing the XploRe language in Yxilon!

Modularity Each part of the software strictly separated.

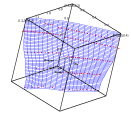
Extensibility driven by modularity single components can be tailored individually

...

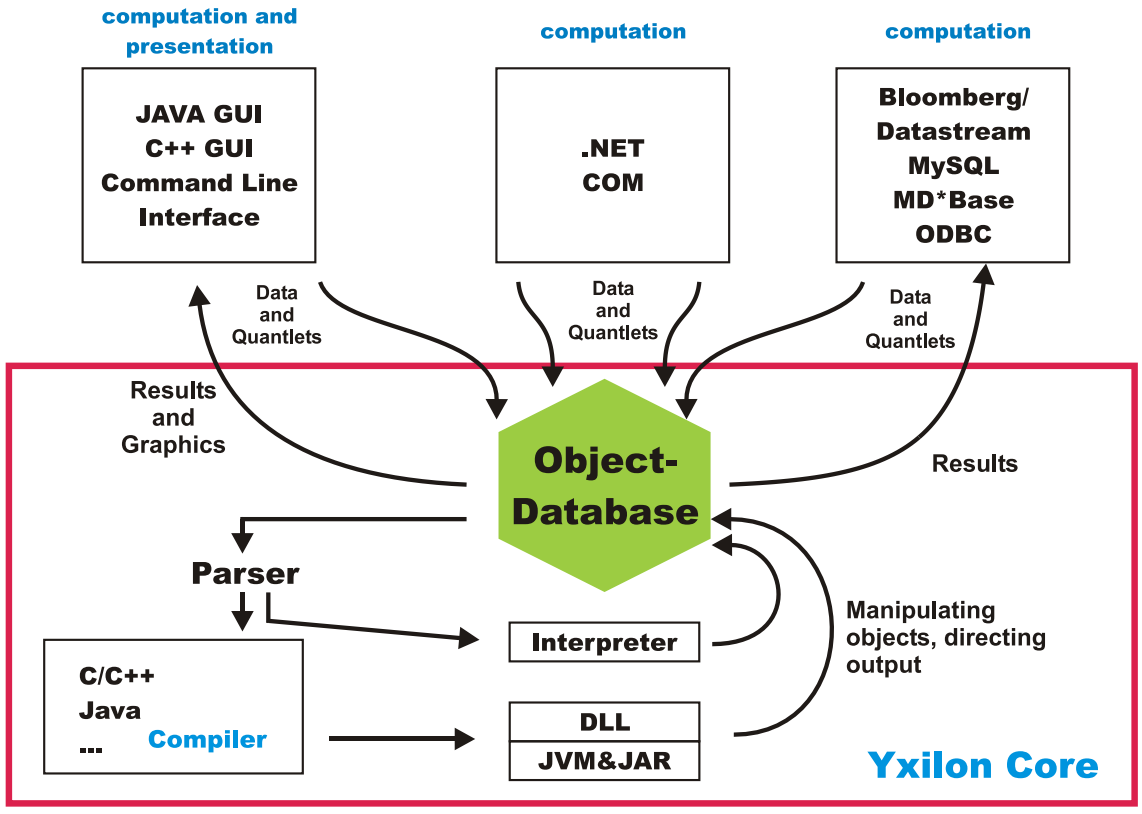


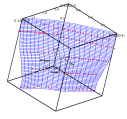
Goal: Vertical Integration

	Scenario 1	Scenario 2	Scenario 3
Organisation	Excel	webservice data	LaTeX
Analysis	stat. engine	computing engine	Applet
Presentation	Excel	trading signal	browser



Internal Structure of Yxilon





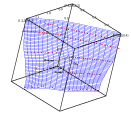
Internal Structure of Yxilon

Object Database

- acts as central repository for data, quantlet sources
- offers connections to databases and user clients
- controls parsing/compilation and execution of code

Parser

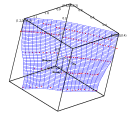
- analyzes quantlets for lexical and semantic errors
- generates JAVA/C++ code



Internal Structure of Yxilon

Interpreter
- compilation→debugging→Execution
- for non time-critical applications interpreter

Clients
- divide into user and framework/database clients
- non-/graphical user clients as human interfaces
- framework/ database clients to store/retrieve data and methods and offer connectivity to CORBA, .NET, COM



Stages in the Development

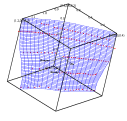
Two main stages:

1st stage

- specification of XploRe language core
- implementation of code generators for Java and C++
- development of platform-independent graphical user interface

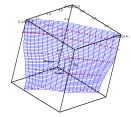
2nd stage

- specification and implementation of persistent object database
- implementation of code generators for Fortran and COM
- development of specialized user interfaces



Usability

Learnability	How easy can first-time users accomplish basic tasks?
Efficiency	How quickly can users perform tasks once they know how to use the software?
Memorability	When users return to the design after a period of not using it, how easily can they reestablish proficiency?
Errors	How many (severe) errors do users make, how is the prevention of errors supported?
Satisfaction	How pleasant is it to use the software?

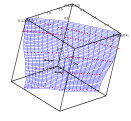


Golden Rules of Interface Design I

consistency similar reactions from the software in similar situations expected by software

shortcuts and feedback beginners need guidance, experienced users just want to finish

closed actions Actions executed by the user need well-defined start and end points.

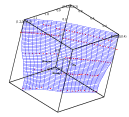


Golden Rules of Interface Design II

error handling Error messages should be short and informative, "There was an error" is not sufficient.

loss of control Users prefer to act, not to react when working with software

limited short term memory Humans only able store just a few items in short term memory. User interface should reflect this



Conclusion and Invitation

- Yxilon is our proposal for a modern statistics package
- Sourcecode and binaries at `http://www.quantlet.org`
- feedback from users and programmers highly appreciated
- Join the Yxilon project!