



Adapting to the technological landscape

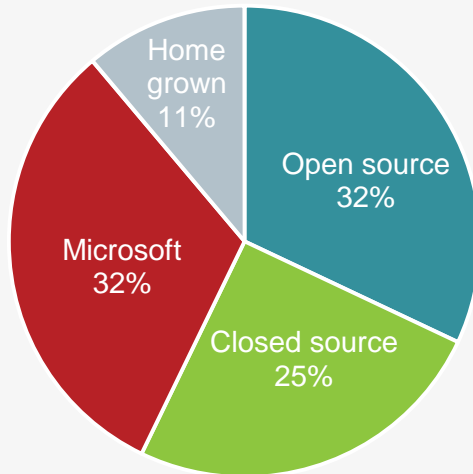
06/18/2018

Conrad Chen, AVP Integrated Analytics

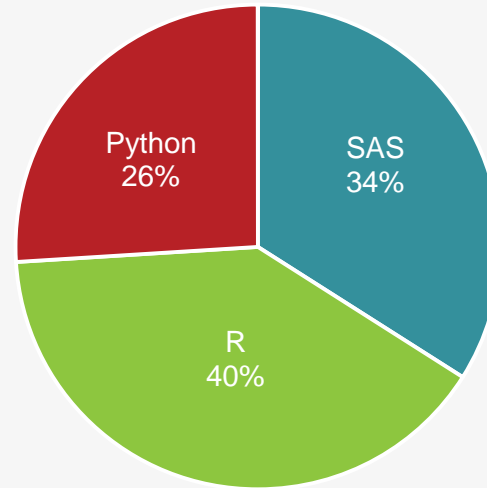
- Adoption of new tools
- Brief word on computing
- Analytic strategy
- More data and more factors
- Maturity of open source tools
 - R language/ RStudio
 - R Markdown/ R Notebook
 - Shiny app

What modeling software do you use?

Life actuaries
2015 SOA survey



Quantitative professionals
2017 Burtch Works survey



A brief history of adopting new tools

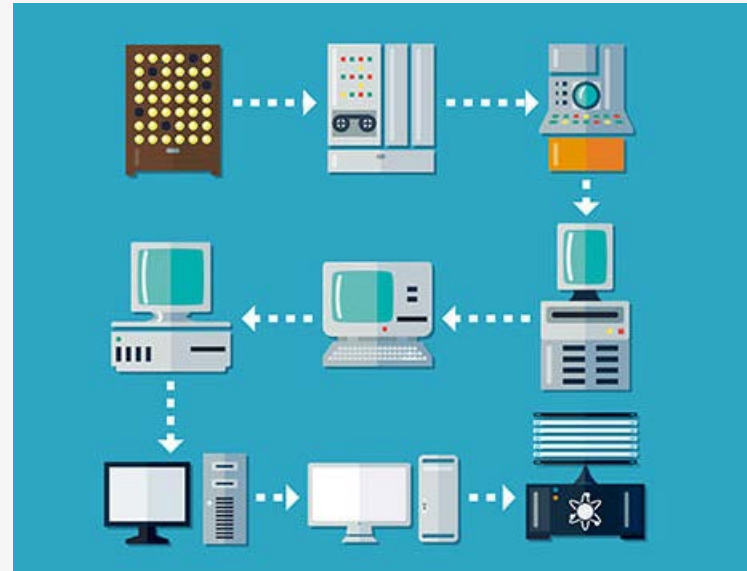
- Tabulation machine 1890 U.S. Census
 - The Census Bureau sought a more efficient method to process and tabulate data than hand counting—the 1880 census did not finish until 1887.

- Pocket calculators in the mid-1970s
 - Use of slide rule declined with technical workers in the 1960s because of early computers, and eventually disappeared with the emergence of inexpensive scientific calculators



Evolution of computing

- Mainframe computers
- Virtualization
- PCs
- Client/server
- Web based
- Cloud computing





- Radically shift how we do business using data, technology, and advanced analytic techniques including machine learning and artificial intelligence
- Data science solutions must be engineered into platforms for application to core processes
- Bring together to generate value

Many sources of information

8. What data sources did your company use to develop the predictive model?

Table A.13 shows the sources companies used in developing their PA model.

Table A.13 – Sources of Information

Data Source	Marketing	Underwriting	Post-Issue Mgmt.
Vendor data (e.g., lab, credit bureau, etc.)	17	3	4
Financial data	16	4	4
Lifestyle data	13	4	3
Internal experience study	12	4	2
Application	12	3	3
Internal data from other lines of business (e.g., P & C)	10	0	5
Actual physical measurements	9	3	0
External consulting data	8	2	1
MIB checking service	5	5	0
MVR	5	5	0
Prescription history	5	4	0
Personal history report	5	3	0
Criminal history	4	4	0
Credit report or bankruptcy	4	3	1
Reinsurer data	4	0	0
MIB Insurance Activity Index	3	6	0
Special questionnaire (e.g. older age, aviation)	3	0	0
Other internal data	2	0	1
Online questionnaire	1	0	0
Social media	1	0	0
Health-related technology (e.g. wearable, smart phone)	0	0	0
Other	0	0	2

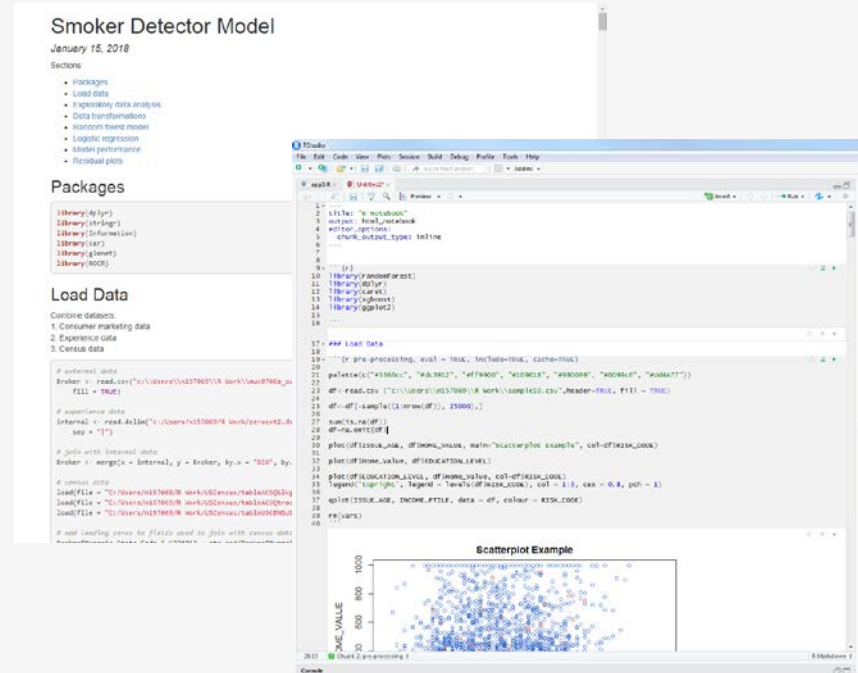
2017 SOA report on predictive analytics and accelerated UW

- Lots of data available
- Incorporate into actuarial analysis
- Assess correlation and interactions in data

- With open source tools you don't need to involve procurement
- Tools that are closed licensed and un-scriptable limit the scope of problems you can solve
- Excel is a great tool in many situations, but know when other tools are better suited
- Build skills in portable and scalable technologies
- Growth in free, accurate, and valuable open source help platforms such as Stack Overflow and Cross Validated
- The R Language appeared in the early '90s and is used widely for statistical computing with a command line interface
- Rstudio was released in 2011 and is a free graphical front-end to facilitate R programming

- Need reproducibility in order to have flexibility to make changes in analysis, e.g., to produce new plots or add/remove variables
- Typically there are many iterations in a project
- Challenge to document and keep track
- Open source tools R Shiny and R markdown to help with above tasks

- R markdown is a package to write results of analyses into common formats (i.e., HTML, PDF, and Word)
- It is a method to create quick and reproducible reporting from R
- R Notebook is an R Markdown document with R code that can be executed interactively
- The notebook's interaction with R makes it a good choice when authoring the R Markdown document and iterating on code



The screenshot displays an R Notebook titled "Smoker Detector Model" dated January 15, 2018. The interface is divided into several sections: "Sections" (listing packages, load data, data transformations, random forest model, logistic regression, model performance, and residual plots), "Packages" (listing libraries like rmarkdown, stringr, furrr, car, gplots, and rOCR), "Load Data" (containing code to read data from a database and perform preprocessing), and "Code" (the main execution area). The code includes library calls, data loading, preprocessing steps like creating a factor variable for smoking status, and plotting functions. A scatterplot titled "Scatterplot Example" is shown at the bottom right, with "SMK_VALUE" on the y-axis (ranging from 0 to 1000) and "EDUCATION_LEVEL" on the x-axis (ranging from 0.0 to 1.0). The plot shows a dense cloud of blue data points.

- Incorporating visualization to the analysis can help deepen understanding of data and convey insights
- Shiny is an R package to build web applications using R
- A Shiny app enables users to visualize and interact with data, where R code is executed on the backend.
- Your analysis is transformed into interactive web application
- Can be used with R markdown

