

# 5.5 Selection of Software Tools

## *Dimensions of Software*

- 1) Statistical capabilities
- 2) Data Mining
- 3) Simulation
- 4) Optimization
- 5) Visualization / Reporting

Others: User-friendliness, costs (cost of license vs. open-source), maintenance, transparency,...

# 5.5 Selection of Software Tools

Software	Statistical	Data Mining	Simulation	Optimisation	Visualisation
Excel	Medium	Medium	Low	Low	High
Access	Medium	Medium	Low	Low	Low
SQL	Low	Medium	Low	Low	Low
R	High	High	Medium	Medium	Low
KNIME	Medium	High	Low	Low	Medium
WEKA	Medium	High	Low	Low	High
SPSS	High	High	Low	Low	Medium
SAS	High	Medium	Medium	High	Medium
Rapidminer	Medium	High	Low	Low	Medium
Matlab	Medium	Medium	High	High	Medium

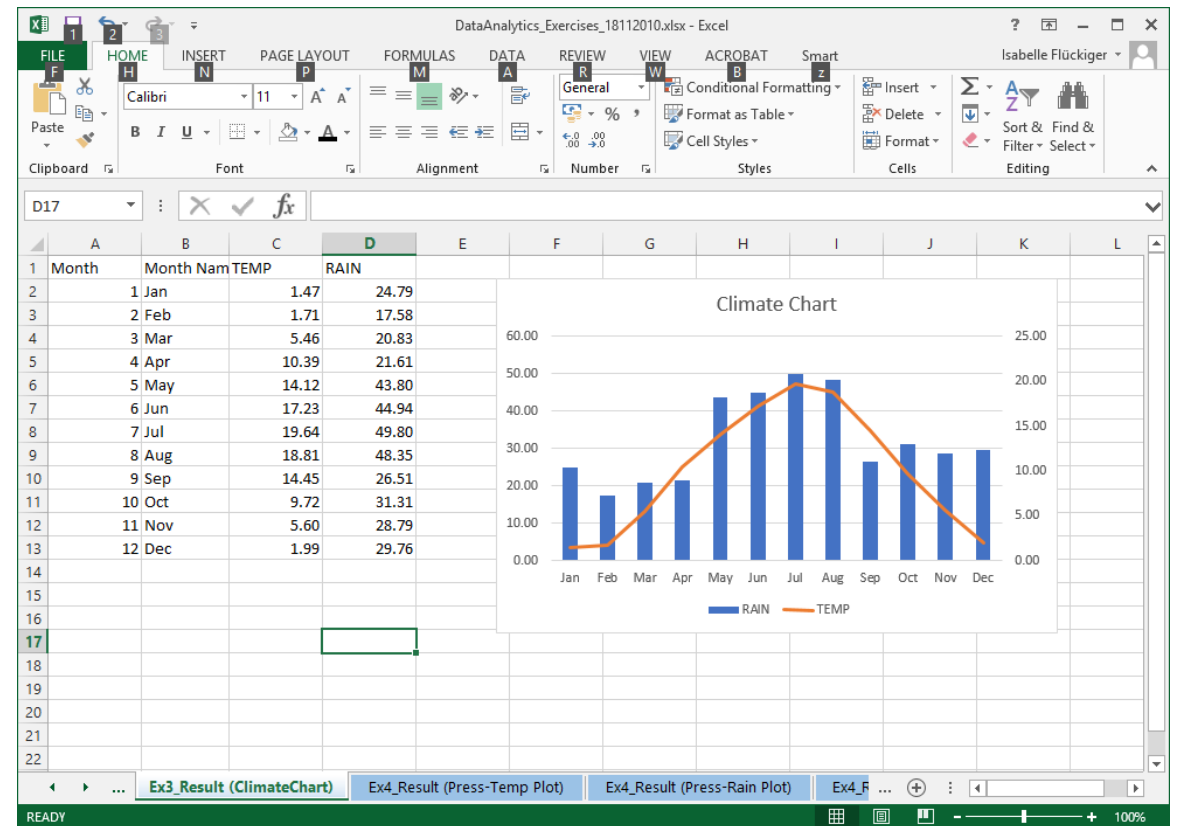
# 5.5 Selection of Software Tools

Software	Statistical	Data Mining	Simulation	Optimisation	Visualisation
@Risk	Medium	Low	High	Medium	High
Crystal Ball	Medium	Low	High	Low	Medium
Python (incl pandas)	High	High	Medium	Medium	Low
Julia	High	High	Medium	Medium	Low
Vensim	Low	Low	High	Medium	Medium
PowerSim	Low	Low	High	Medium	Medium
Anylogic	Low	Low	High	Medium	Low
Qlik (QlikView, Qlik Sense)	Low	Medium	Low	Low	High
Tableau	Low	Medium	Low	Low	High

# 5.5 Selection of Software Tools

## *Microsoft Excel*

- *Microsoft Office Application*
- *“easy to use” spreadsheet tool*
- *It contains calculations, graphic tools and pivot tables*
- *Additionally, macro programming is integrated, called Visual Basic for Applications (VBA)*
- *Advantage: all companies have this*
- *There are add-on like @Risk such that even more statistical analyses and stochastic simulation can be performed*



# 5.5 Selection of Software Tools

R (<https://www.r-project.org/>)

- “R is a language and environment for statistical computing and graphics.”
- Open Source (GNU licenses)
- You have many, many contributed packages... ([https://cran.r-project.org/web/packages/available\\_packages\\_by\\_name.html](https://cran.r-project.org/web/packages/available_packages_by_name.html))



## The R Project for Statistical Computing

[Home]

### Download

CRAN

### R Project

About R  
Contributors  
What's New?  
Mailing Lists  
Bug Tracking  
Conferences  
Search

### R Foundation

Foundation  
Board  
Members  
Donors  
Donate

### Documentation

Manuals  
FAQs  
The R Journal  
Books  
Certification  
Other

### Links

Bioconductor  
Related

## Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

## News

- **R version 3.2.3 (Wooden Christmas-Tree) prerelease versions** will appear starting Monday 2015-11-30. Final release is scheduled for Thursday 2015-12-10.
- **R version 3.2.2 (Fire Safety)** has been released on 2015-08-14.
- **The R Journal Volume 7/1** is available.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.
- **useR! 2015**, took place at the University of Aalborg, Denmark, June 30 - July 3, 2015.
- **useR! 2014**, took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.

# 5.5 Selection of Software Tools

*Example:* R packages related to descriptive, predictive and prescriptive analytics:

R PACKAGES					
caret	kernlab	Gmodels	mlr		
abind	dichromat	igraph	pkgmaker	ROracle	TSP
AER	digest	iplots	playwith	roxygen2	urca
and	doParallel	irlba	pls	rpart	utils
ArDec	dplyr	iterators	plyr	rpart.plot	vcd
arm	dse	jsonlite	png	rpartOrdinal	whisker
			princomp		
arules	dtw	KernSmooth	pROC	RpgSQL	WriteXLS
	e1071	kknn			
arulesViz	eptools	knitr	proto	RPostgreSQL	XLConnect
assertthat	EMCC	labeling	qcc	RQuantLib	xlsReadWrite
ast	evaluate	lattice	quantmod	RSNNS	xlsx
base	extremevalues	lazyeval	quantreg	RSQLite	xlsx
bayesclust	ff	lme4	R2HTML	rstudioapi	xml2
bclust	ffbase	lmtree	R2PPT	sampleSelection	xtable
BH	foreach	lubridate	R6	sandwich	xts
biclust	forecast	magrittr	randomForest	scales	yaml
bitops	foreign	maptools	rCharts	scatterplot3d	zoo
boot	formatR	markdown	RColorBrewer	seriation	
bootstrap					
brew	fpc	MASS	Rcpp	shiny	
cairoDevice	gbm	Matrix	RcppEigen	sp	
car	gcExplorer	mclust	RCurl	SPARQL	
caTools	gclus	memisc	registry	SparseM	
chron	gdata	memoise	regr	spatial	
class	ggplot2	methods	regro	splines	
clue	ggvis	mgcv	rEMM	sqldf	
clues	git2r	mime	reshape	stasts	
cluster	googleVis	minqa	reshape2	statnet	
clusterGeneration	gplots	mlogit	response	stats	
clusterSim	graphics	Modalclust	rggobi	stats4	
clustsig	grDevices	moments	rgl	stringi	
clv	grid	multcomp	RGtk2	stringr	
clValid	gridBase	munsell	rJava	survival	
coda	gtable	mvoutlier	RJDBC	teltk	
codetools	gtools	neuralnet	RJSONIO	termstrc	
colorspace	gWidgets	nlme	Rlof	tidyr	
compiler	gWidgetsRGtk2	nloptr	rmarkdown	timsac	
curl	highr	NMF	Rmongo	tm	
data.table	htmltools	nnet	RMYSQL	tools	
datasets	httpuv	nortest	rngtools	translations	
DBI	httr	outliers	ROCR	tree	
Design	hybridHclust	parallel	RODBC	truncreg	
devtools	iCluster	pbrktest	RODM	TSclust	

# 5.5 Selection of Software Tools

*KNIME* (<https://www.knime.org/>)

- Konstanz Information Miner
- Data mining, machine learning, reporting and reporting / visualization
- R and WEKA integration
- Java based
- Open Source

The image displays two screenshots of the KNIME website. The top screenshot shows the KNIME Spring Summit 2016 announcement, featuring a silhouette of the Berlin skyline and a 'KNIME Spring Summit 2016' banner. Below the banner are sections for 'The open platform for your data' and 'Adding value'. The bottom screenshot shows the 'Products' page, which lists various KNIME products and extensions, including KNIME Analytics Platform, KNIME Personal Productivity, KNIME Partner Productivity, KNIME TeamSpace, KNIME Server Life, KNIME Server, KNIME Big Data Extension, KNIME Cluster Execution, and KNIME Product Matrix. A central diagram shows the KNIME Analytics Platform at the core, surrounded by various extensions like KNIME TeamSpace, KNIME Server Life, KNIME Webfrontal, KNIME Big Data Extension, KNIME Cluster Extension, KNIME Personal Productivity, KNIME Partner Productivity, and Community & Partner Extensions.

# 5.5 Selection of Software Tools

*KNIME*

(<https://www.knime.org/>)

- Graphical interface
- Works with nodes and data flows

The screenshot displays the KNIME software interface. On the left, there is a 'Node Repository' pane showing various data processing nodes. The main workspace contains a data flow diagram with nodes such as 'Database Reader', 'Image Reader', 'Image Cropper', 'File Reader', 'Color Manager', 'Interactive Table', 'Interactive Segmentation View', 'File Reader', 'File Writer', and 'CSV Writer'. A 'k-Means' node is highlighted, and its configuration dialog is open on the right. The dialog includes a description of the algorithm, 'Dialog Options' for 'number of clusters' and 'size number of iterations', and 'Ports' for 'Input Ports' and 'Output Ports'. At the bottom, a 'Console' window shows the execution log, including the KNIME version (v3.7.0.0000000) and the file path for the log file.

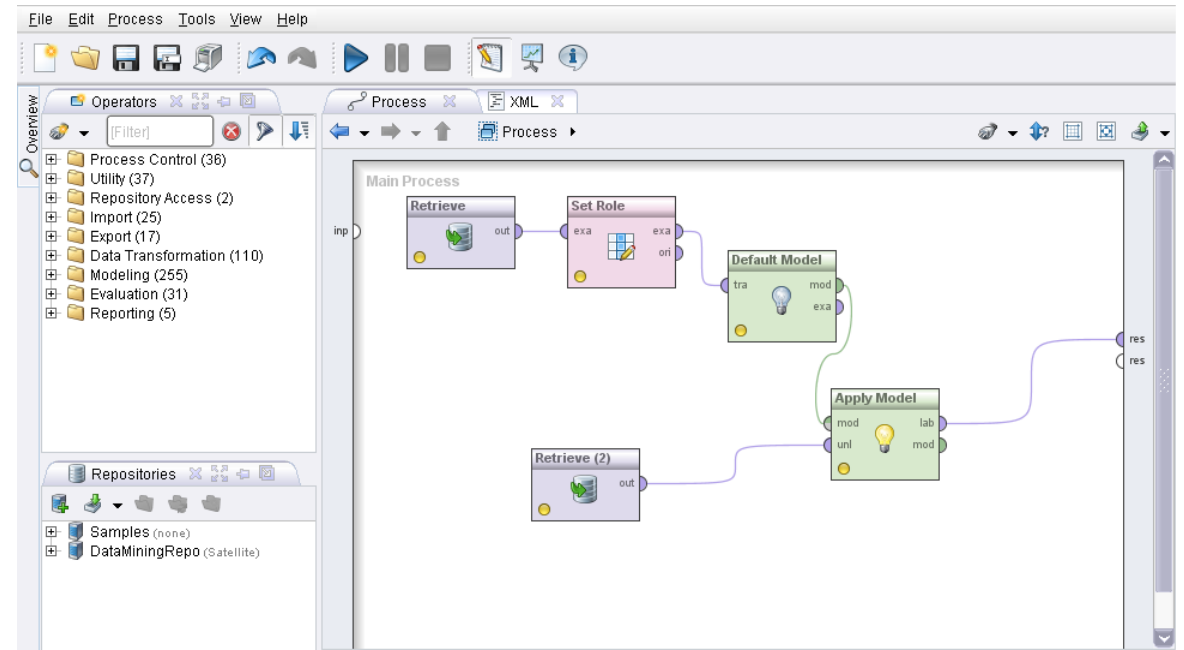


# 5.5 Selection of Software Tools

## *Rapidminer*

(<https://rapidminer.com/>)

- Machine learning, data mining, text mining, web mining, predictive analytics and business analytics
- Community Edition for free, licensing for more comprehensive versions



# 5.5 Selection of Software Tools

## WEKA

(<http://www.cs.waikato.ac.nz/~ml/weka/index.html>)

- Waikato Environment for Knowledge Analysis
- Data preprocessing, regression, clustering, classification, feature / attribute selection, and visualization
- Open Source



Machine Learning Group at the University of Waikato

[Project](#)   [Software](#)   [Book](#)   [Publications](#)   [People](#)   [Related](#)

## Weka 3: Data Mining Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like **this**, and the bird sounds like **this**.

Weka is open source software issued under the **GNU General Public License**.

Yes, it is possible to apply Weka to **big data**!

**Data Mining with Weka** is a 5 week MOOC, which was held first in late 2013. Check out the **MOOC site** for video lectures and details on how to enrol into this course and a new, advanced Weka course.

### Getting started

- Requirements
- Download
- Documentation
- FAQ
- Getting Help

### Further information

- Citing Weka
- Datasets
- Related Projects
- Miscellaneous Code
- Other Literature

### Developers

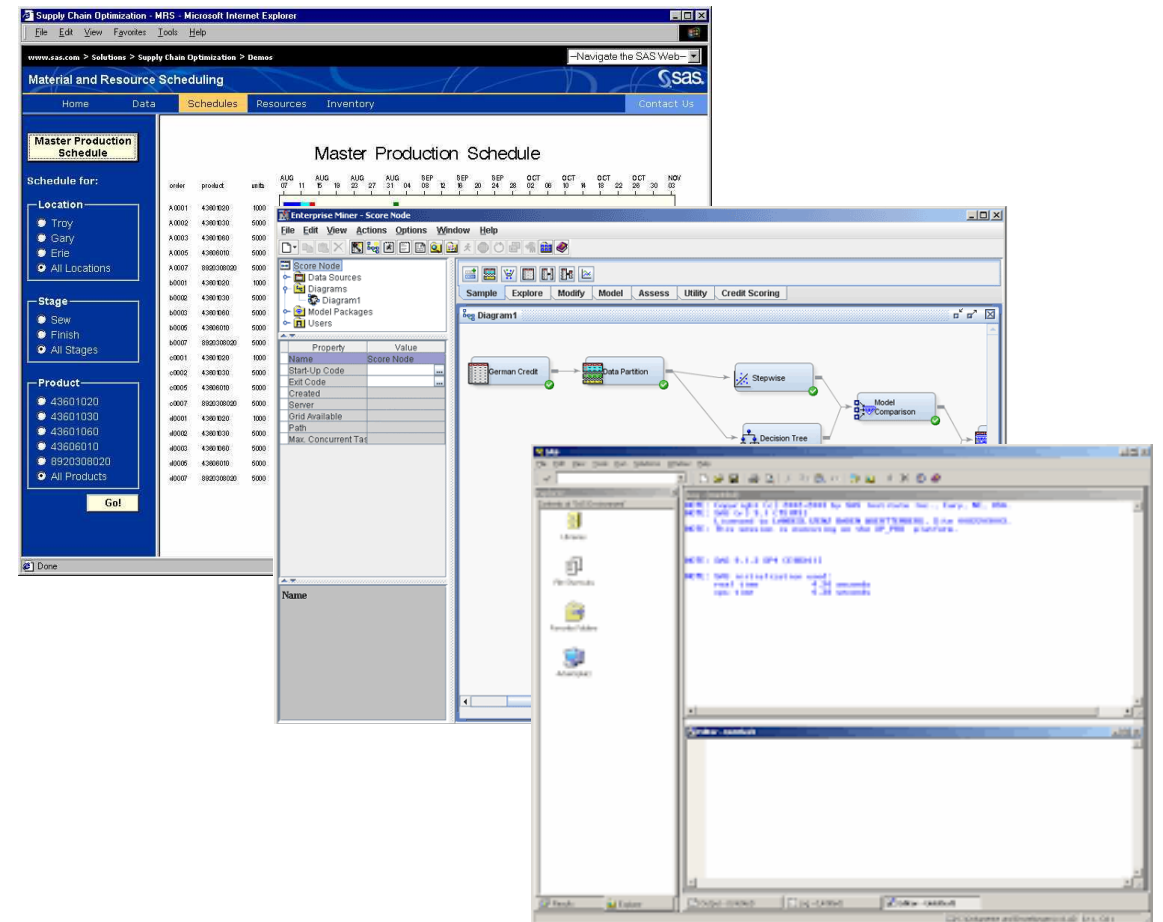
- Development
- History
- Subversion
- Contributors

# 5.5 Selection of Software Tools

SAS

([http://www.sas.com/en\\_gb/home.html](http://www.sas.com/en_gb/home.html))

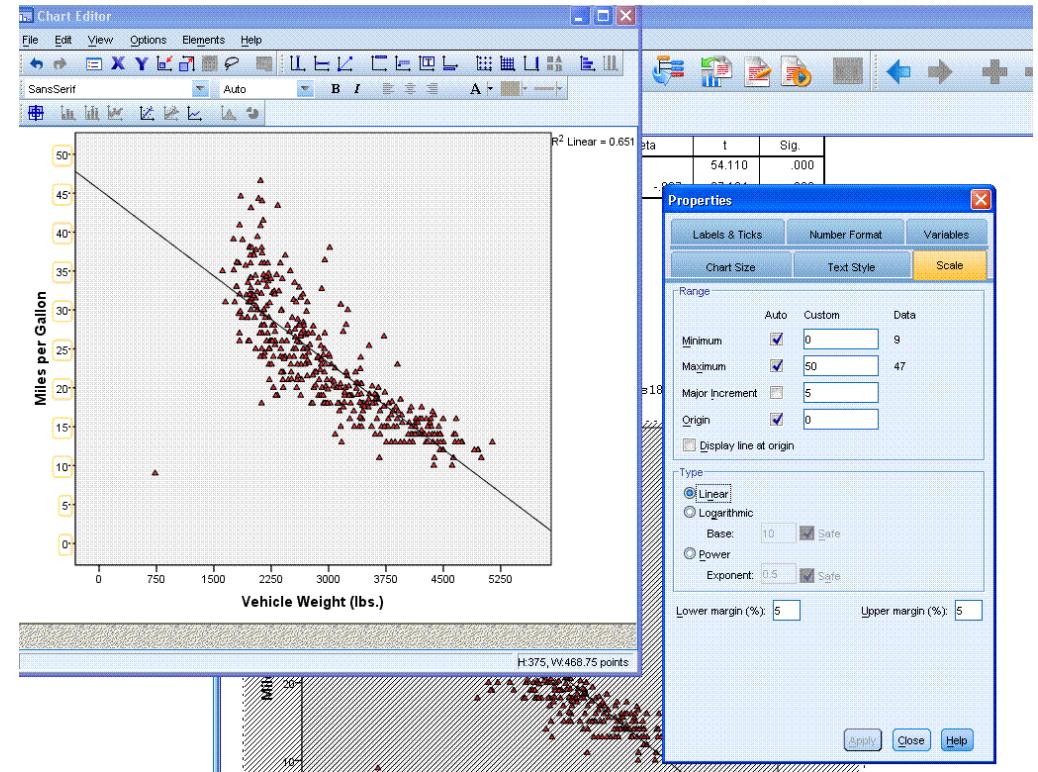
- Statistical programming language for data analytics, data management, business intelligence, risk management, supply chain management
- Various industry specific solutions
- Licensing model
- Very often used in industries to proceed large amount of data



# 5.5 Selection of Software Tools

*IBM SPSS Statistics* (<http://www-03.ibm.com/software/products/en/spss-statistics>)

- Professional vendor suite
- Licensing system i.e. for each module one needs a license and has to pay
- “addresses the entire analytical process, from planning and data collection to analysis, reporting and deployment”



# 5.5 Selection of Software Tools

IBM Software | Products | Business analytics | Predictive analytics

## SPSS Statistics

Family | Downloads

Solve difficult business and research challenges with data analysis

IBM SPSS Statistics is an integrated family of products that helps to address the entire analytical process, from planning and data collection to analysis, reporting and deployment. With more than a dozen fully integrated modules to choose from, you can find the specialized capabilities you need to increase revenue, outperform competitors and make better decisions.

Product editions:

- [SPSS Advanced Statistics](#): Analyze complex relationships using powerful univariate and multivariate analysis. | Product Features | Pricing support
- [SPSS Amos](#): Build attitudinal and behavioral models that realistically reflect complex relationships. | Product Features | Pricing support
- [SPSS Bootstrapping](#): Test the stability of your analytical models and procedures. | Product Features | Pricing support
- [SPSS Categories](#): Predict outcomes and reveal relationships in categorical data. | Product Features | Pricing support
- [SPSS Complex Samples](#): Compute statistics correctly for complex sampling. | Product Features | Pricing support
- [SPSS Conjoint](#): Understand consumer preferences for more effective market research. | Product Features | Pricing support
- [SPSS Custom Tables](#): Summarize SPSS Statistics data in different ways for different audiences. | Product Features | Pricing support

Not in United Kingdom?  
Select another country.

Considering a purchase?

Contact IBM

- E-mail IBM
- Or call us at: 01483 718 528  
Priority code: SPSS

**Product support**

- Brochure: IBM SPSS Statistics Editions
- White Paper: The power of IBM SPSS Statistics and R together
- White Paper: Better decision making under uncertain conditions

[SPSS Data Preparation](#): Perform faster, more accurate data preparation and validation. | Product Features | Pricing support

[SPSS Decision Trees](#): Explain categorical results in a clear, intuitive manner. | Product Features | Pricing support

[SPSS Direct Marketing](#): Improve your campaigns and maximize marketing ROI. | Product Features | Pricing support

[SPSS Exact Tests](#): Analyze small datasets or those with rare occurrences more accurately. | Product Features | Pricing support

[SPSS Forecasting](#): Build sophisticated time-series forecasts regardless of your skill level. | Product Features | Pricing support

[SPSS Missing Values](#): Estimate missing data so you can draw more valid conclusions. | Product Features | Pricing support

[SPSS Neural Networks](#): Find more complex relationships in your data. | Product Features | Pricing support

[SPSS Regression](#): Classify your data into groups and predict categorical outcomes. | Product Features | Pricing support

[SPSS SamplePower](#): Find the right sample size for your research project. | Product Features | Pricing support

[SPSS Statistics Base](#): Get essential tools for every step of the analytical process. | Product Features | Pricing support

[SPSS Statistics Campus](#): Provide unlimited numbers of users with access to analytics. | Product Features support

[SPSS Statistics Developer](#): Create custom features for SPSS Statistics using R and Python.

[SPSS Statistics Faculty Pack](#): Incorporate analytics into your undergraduate or graduate curricula. | Product Features support

[SPSS Statistics GradPack](#): Develop advanced analytical skills using this affordable student software package. | Product Features support

[SPSS Statistics Premium](#): An all-in-one edition for enterprise businesses with multiple advanced analytics requirements. | Product Features | Pricing support

[SPSS Statistics Professional](#): Tools to address the challenges of the entire analytic lifecycle. | Product Features | Pricing support

[SPSS Statistics Programmability Extension](#): Add procedures written in Python, R, Java and Microsoft Visual Basic.NET. | Product Features support

[SPSS Statistics Standard](#): Fundamental analytical capabilities for a range of business and research questions | Product Features | Pricing support

[SPSS Statistics for Linux on System z](#): Combine analytical power with the security and stability of System z. Features

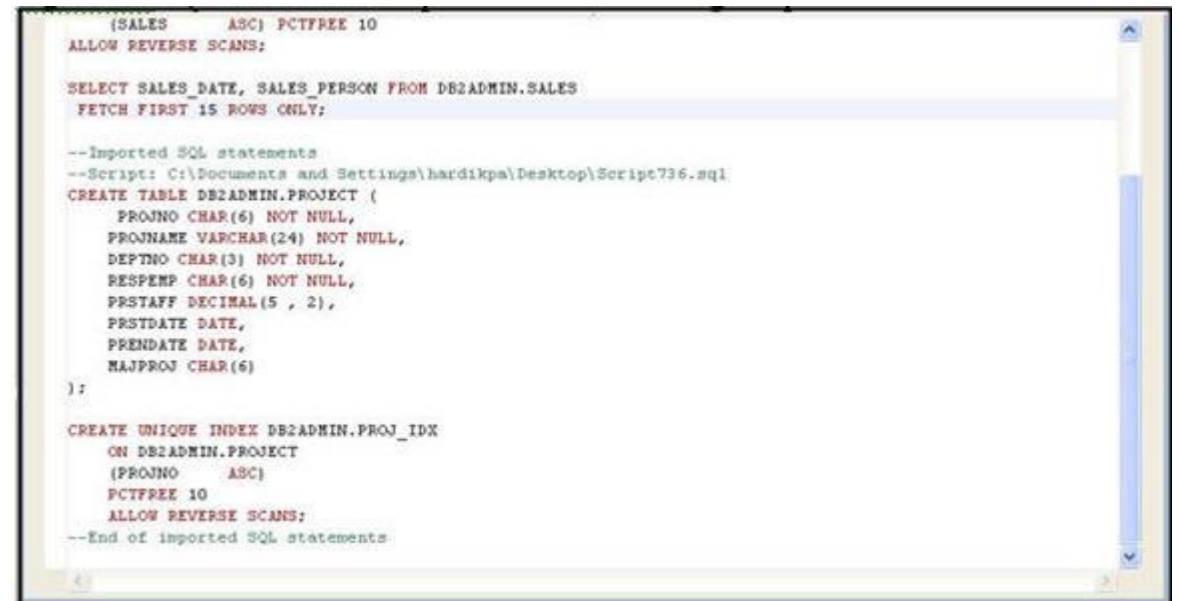
[SPSS Text Analytics for Surveys](#): Turn open text survey responses into quantifiable data through programmatic text analysis. | Product Features | Pricing support

[SPSS Visualization Designer](#): Create and share customized data visualizations, from simple charts to advanced graphics.

# 5.5 Selection of Software Tools

## *SQL*

- Structured Query Language
- Language for managing structure data held typically called “data in a relational database management system”
- Advanced “filtering” tool, i.e. query data
- Used in all database types like Oracle, SAP, IBM DB, Microsoft and so on



```
(SALES ASC) PCTFREE 10
ALLOW REVERSE SCANS:

SELECT SALES_DATE, SALES_PERSON FROM DB2ADMIN.SALES
FETCH FIRST 15 ROWS ONLY;

--Imported SQL statements
--Script: C:\Documents and Settings\hardikpa\Desktop\Script736.sql
CREATE TABLE DB2ADMIN.PROJECT (
  PROJNO CHAR(6) NOT NULL,
  PROJNAME VARCHAR(24) NOT NULL,
  DEPTNO CHAR(3) NOT NULL,
  RESPENS CHAR(6) NOT NULL,
  PRSTAFF DECIMAL(5, 2),
  PRSTDATE DATE,
  PRENDATE DATE,
  MAJPROJ CHAR(6)
);

CREATE UNIQUE INDEX DB2ADMIN.PROJ_IDX
ON DB2ADMIN.PROJECT
(PROJNO ASC)
PCTFREE 10
ALLOW REVERSE SCANS:
--End of imported SQL statements
```

# 5.5 Selection of Software Tools

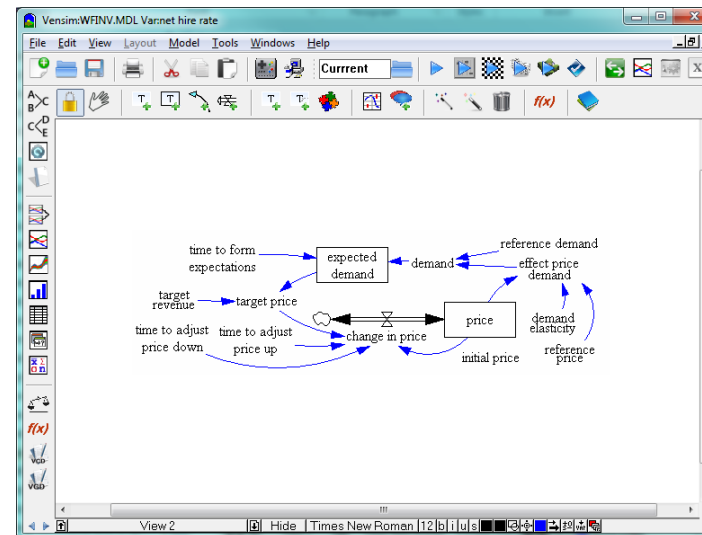
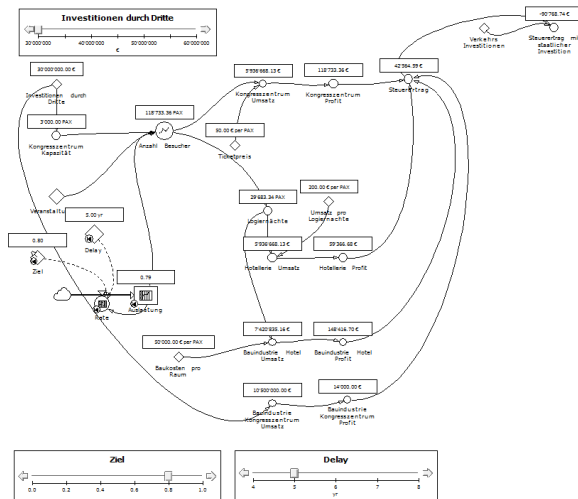
And there are many more programming languages or frameworks used like

- *Apache Hadoop*: open-source software framework in Java “for distributed storage and distributed processing of very large data sets”
- *.NET Framework*: it is a software framework developed by Microsoft and includes user interfaces, data access management, database connectivity, cryptography, web application development, numeric algorithms and network communications. E.g C# is one part of it.
- *Java*
- *Julia* (<http://julialang.org/>): high-level dynamic programming language used for scientific computing, machine learning, data mining, large-scale linear algebra, distributed and parallel computing; it is a very efficient and effective language.
- and so on

# 5.5 Selection of Software Tools

## *System Dynamics Software*

- PowerSim (<http://www.powersim.com/>)
- Vensim (<http://vensim.com/>): free edition for educational purpose





# Excuse System Dynamics

*What is System Dynamics?*

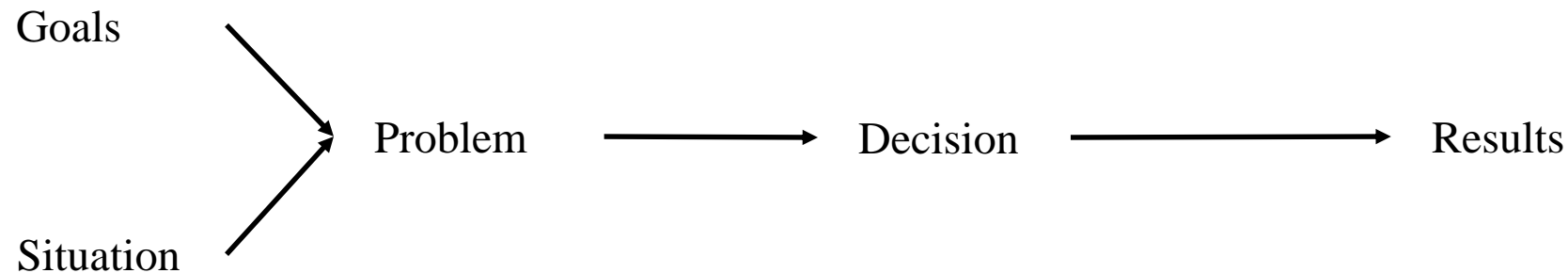
System Dynamics (SD) is a methodology for framing, modelling and understanding the dynamic of complex system.

It is used for the understanding of non-linear behavior of systems.

It is finally a tool to support us in system thinking.

# Excuse System Dynamics

*Typical thinking:*

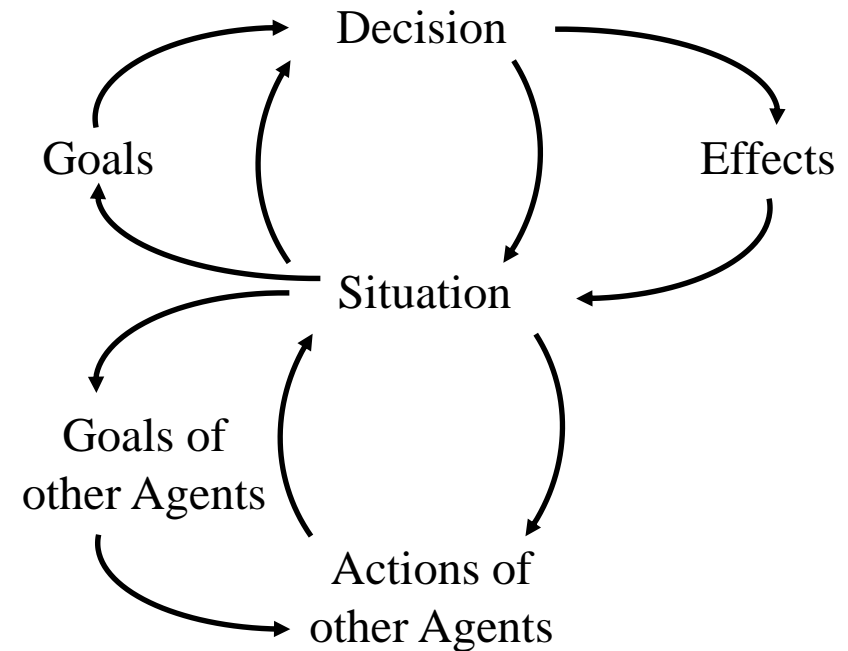


In the context of risks this is called “root – cause” analysis.

# Excuse System Dynamics

*Alternative thinking:*

- move away from looking at isolated events and their causes
- start to look at the organization as a system made up of interacting parts



# Excuse System Dynamics

*Why system thinking is important?*

Today's world and the business problem one has to solve are

- Dynamic and evolving over time
- Non-linear
- Have accumulations and delays
- And have typically feedback loops

# Excuse System Dynamics

## *Patterns of Behavior*

First, one is considering patterns of behavior of the variables that characterize the problem and/or situation.

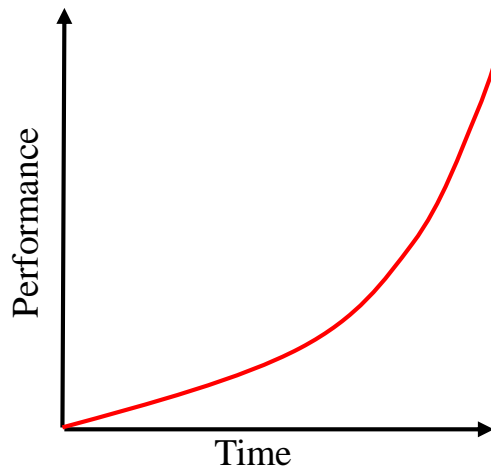
In a business context such variables are:

- Costs
- Sales
- Revenue
- Profit
- Market share
- Risk
- And so on

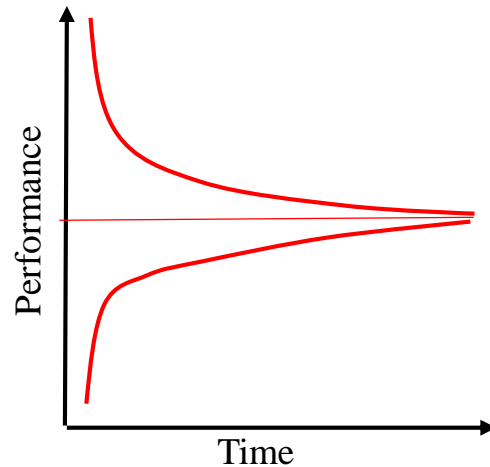
# Excuse System Dynamics

## *Patterns of Behavior*

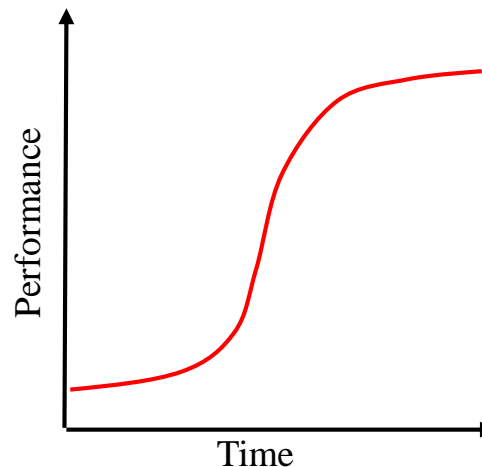
Typical patterns of behavior over time are:



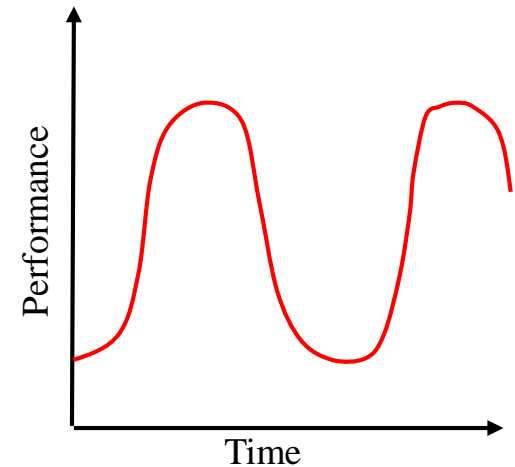
Exponential



Goal-seeking



S-Shape



Oscillation

# Excuse System Dynamics

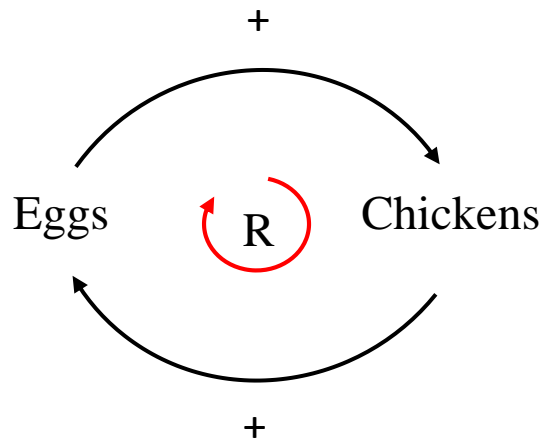
## *Patterns of Behavior*

Common combinations of these four patterns:

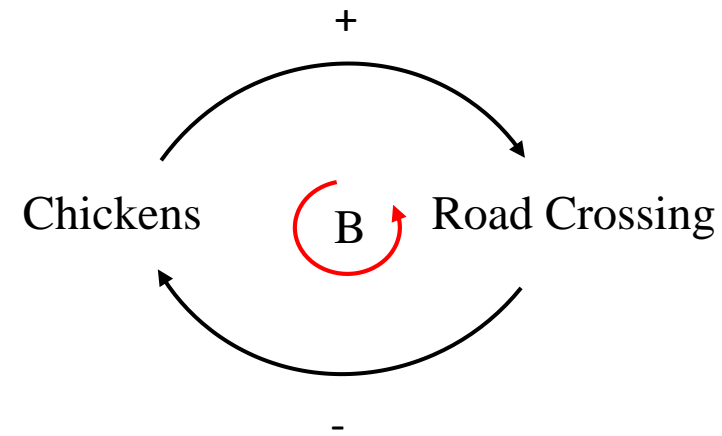
- Exponential growth combined with oscillation
- Goal-seeking behavior combined with an oscillation whose amplitude gradually declines over time
- S-shaped growth combined with an oscillation whose amplitude gradually declines over time

# Excuse System Dynamics

## *Feedback and Causal Loop Diagrams*



Positive Feedback:  
Self-reinforcing (R)



Negative Feedback:  
Self-Correcting / Balancing (B)

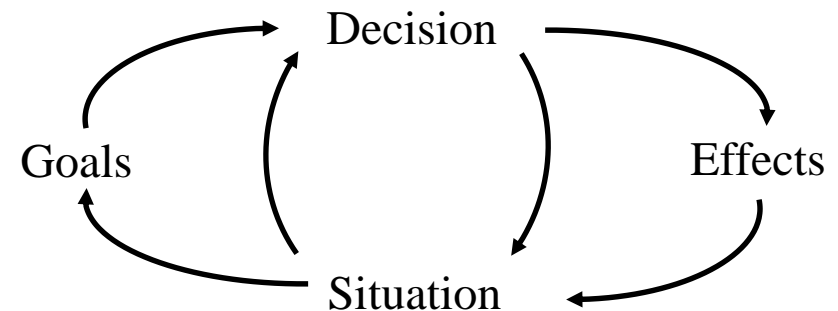


# Excuse System Dynamics

## *Feedback and Causal Loop Diagrams*

*Definition:* When an element of a system indirectly influences itself in the system then, this called a feedback loop or a causal loop.

I.e. a feedback loop is a closed sequence of causes and effects, that is, a closed path of action and information.

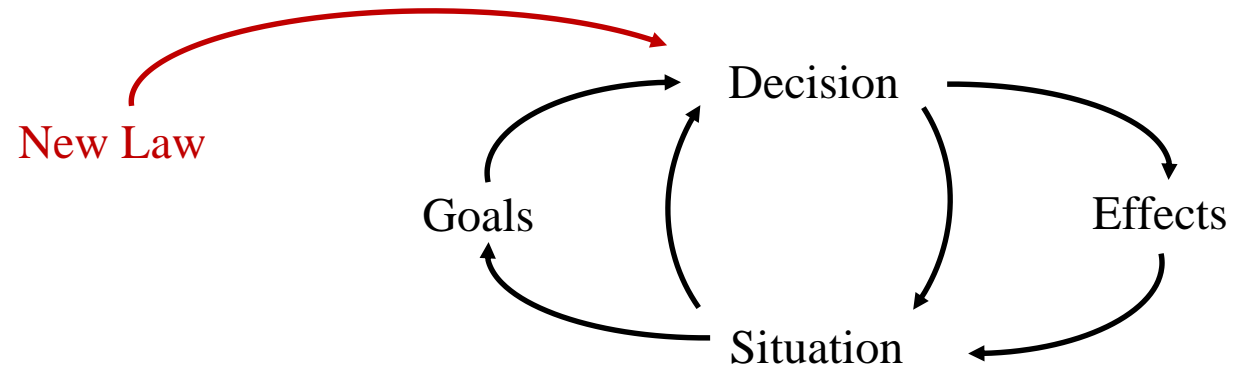


(G. P. Richardson and A. L. Pugh III, Introduction to System Dynamics Modeling with DYNAMO, Productivity Press, Cambridge, Massachusetts, 1981.)

# Excuse System Dynamics

## *Feedback and Causal Loop Diagrams*

*Definition:* An open loop is a linear chain of causes and effects which does not close back on itself



# Excuse System Dynamics

## *How to set up a causal loop diagram (1/2)*

- 1) Think of the elements in a causal loop diagram as variables which can go up or down
- 2) Use nouns or noun phrases to represent the elements e.g. “costs”
- 3) Actions in a causal loop diagram are represented by the arrows
- 4) Ensure that the definition of an element makes it clear which direction is “up” for the variable
- 5) Use an element name for which the positive sense is preferable e.g. “growth”
- 6) Causal links should imply a direction of causation, and not simply a time sequence i.e. “when A increases then B increases.”

# Excuse System Dynamics

## *How to set up a causal loop diagram (2/2)*

- 7) Think about possible unexpected side effects which might occur in addition
- 8) For negative feedback loops, there is a goal
- 9) Include causal loop elements for both the actual value of a variable and the perceived value
- 10) There are often differences between short term and long term consequences of actions, and these may need to be distinguished with different loops
- 11) If a link between two elements needs a lot of explaining, you probably need to add intermediate elements between the two existing elements
- 12) Last but not least: Keep the diagram as simple as possible

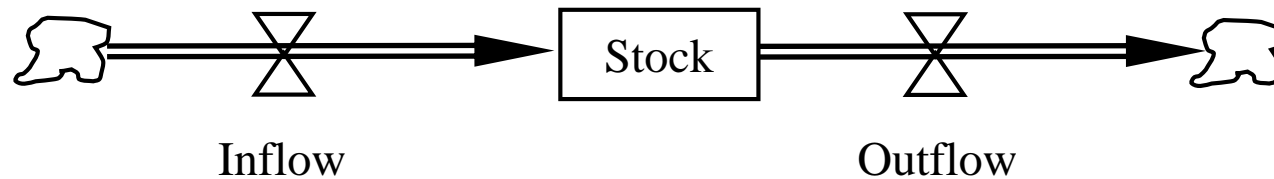
# Excuse System Dynamics

## *Stocks and Flows*



# Excuse System Dynamics

## *Stocks and Flows*

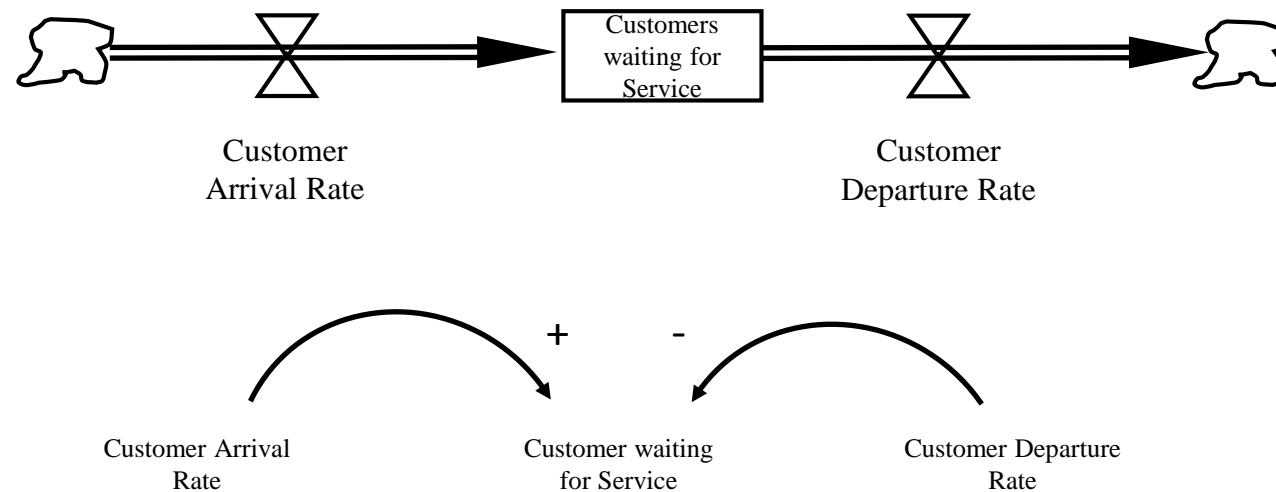


$$Stock(t) = \int_{t_0}^t [Inflow(s) - Outflow(s)] ds + Stock(t_0)$$

$$\frac{d(Stock)}{dt} = Net\ Change\ in\ Stock = Inflow(t) - Outflow(t)$$

# Excuse System Dynamics

## *Stocks and Flows: Example*



# Excuse System Dynamics

## *Auxiliary Variables*

Assume: The rate at which customers can be processed depends on

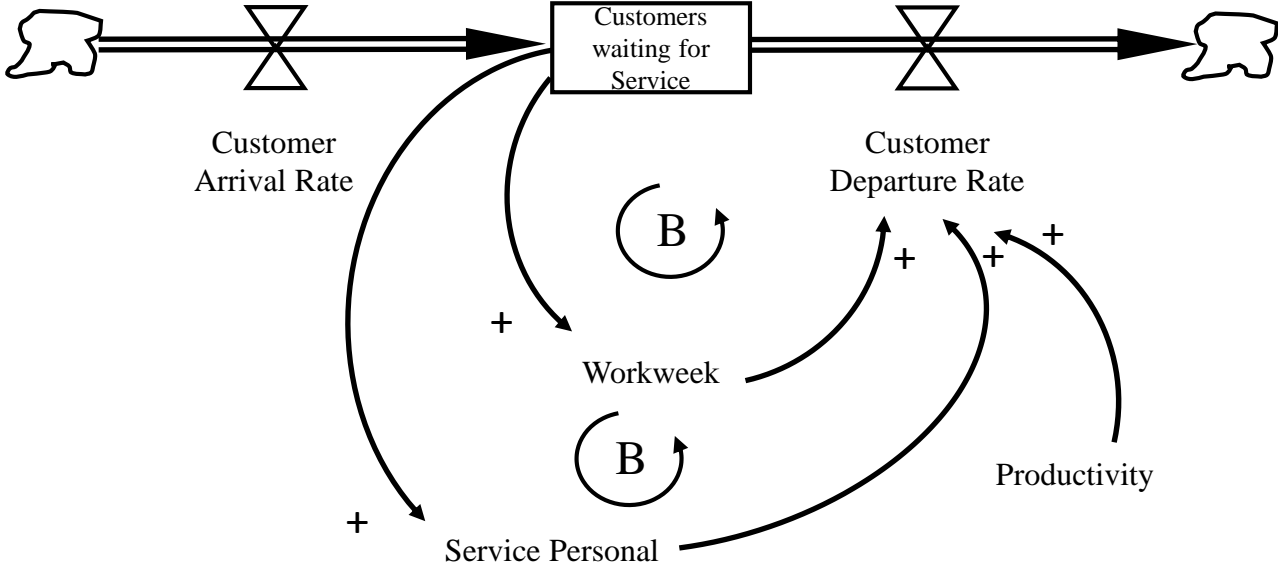
- Number of service personal
- The productivity of the personal (customer processed per hour)
- Numbers of working hours per week

These variables are auxiliary variables which determine the dynamic.



# Excuse System Dynamics

## *Auxiliary Variables*



# Excuse System Dynamics

*Example:*

