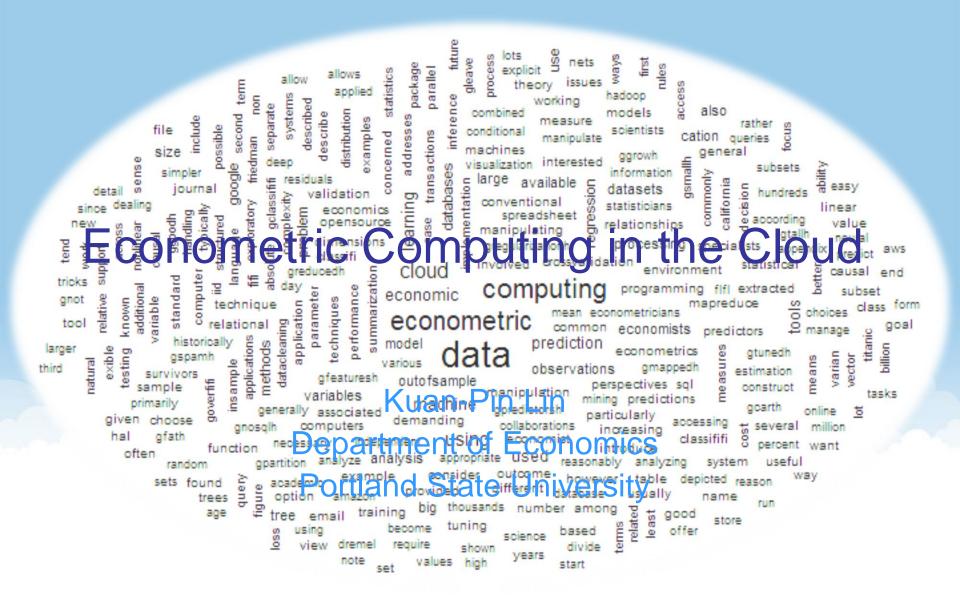
future esn process ways first gleave nets addresses package parallel concerned statistics allows google second term issues access systems hadoop inference working e separate described combined distribution models also describe examples include measure file possible conditional scientists case transactions cation queries friedman manipulate machines ggrowh gsmallh visualization interested databases subsets information residuals
valid
valid
poperation
operation
ope large available odatas datas conventional spreadsheet manipulating gregularizationh production commonly absolute gclassifif california journal datasets decision validation learning fif exploratory statisticians typically linear economics handling dpoodb iid structured opensource according relationships value nonlinear anguage dimensions processing specialists classifi cloud computer better greducedh environment summarization standard application K tricks computing programming fifi extracted additional parameter performance economic gnot relative techniques variable technique mapreduce tools mean econometricians choices econometric datacleaning common economists predictors applications prediction methods model ata larger econometrics varian wector means various insample observations third gmappedh estimation gfeaturesh outofsample perspectives sql sample construct manipulation tasks variables mining predictions primarily machine gcarth gpredictorsh ಕ associated particularly demanding given choose collaborations computers cost increasing gfath hal classififi using economist independent necessary function introduce often gpartition analyze analysis appropriate USEd reasonably random analyzing system outcome consider example query table depicted figure different provided database option usually amazon thousands number among big related training east terms tuning become dremel require divide shown high values start



Econometric Computing

- Econometric Computing: Causal Inference of Economic Data for Decision Making
- Econometrics and Predictive Analytics (Machine Learning, Data Mining)
 - Regression and Classification
 - Model Selection
 - Cross-Validation
 - Prediction

Econometric Computing

- Machine Learning from Econometrics
 - Causal Inference Methods
 - Confounding and Instrumental Variables
 - Regression Discontinuity
 - Difference in Difference
- Advanced Topics
 - Non IID Data (Time Series, Panel Data)
 - High Dimensional Sparse Econometric Models

Using R

- Why R (and RStudio)?
 - Powerful, Popular, and Portable
 - Free (Nothing better than that!)
 - But, with steep learning curve!
- Using R Packages
 - Im(), glm(), plm() for linear modeling
 - tree(), rpart() for machine learning
 - ggplot2() for data visualization
 - parallel() for multi-core and parallel computing

Cloud Computing with R

- Layers of Cloud Computing
 - Cloud Clients ⇔ SaaS, PaaS, laaS
 - laaS: Infrastructure as a Service
 - PaaS: Platform as a Service
 - SaaS: Software as a Service
- Cloud Services Providers
 - Microsoft Azure
 - Google App Engine
 - Amazon EC2, …

Econometric Computing in the Cloud

- A Low Cost Option
 - Amazon Web Services (AWS: Free Tier)
 - Elastic Compute Cloud (EC2)
 - Simple Storage Service (S3)
 - [(\$15 + 10c /gb) / month] after 1st year
 - RStudio Server Amazon Machine Image
 - Implementing R / Rstudio AMI
 - Guide for Dummies (are we?)
 - Louis Aslett's Guide

An Example

- Wine Sales in Vancouver BC
 - Total Weekly Sales of Imported and Domestic Table Wine in Vancouver, BC, Canada from week ending April 4, 2009 to week ending May 28, 2011 (377,228 sales)
 - Irregularly-spaced time series
 - Data Source: <u>American Association of Wine</u>
 <u>Economists</u>

An Example

- Wine Sales in Vancouver BC
 - 377,228 observations of 17 variables in an Excel spreadsheet:

SKU #, Product Long Name, Store Category Major Name, Store Category Sub Name, Store Category Minor Name, Current Display Price, Bottled Location Code, Bottle Location Desc, Domestic/Import Indicator, VQA Indicator, Product Sweetness Code, Product Sweetness Desc, Alcohol Percent, Julian Week No, Week Ending Date, Total Weekly Selling Unit, Total Weekly Volume Litre

An Example

Wine Sales in Vancouver BC

- Data Exploration
 - What = Store Category Minor Name (Red/White)
 - Where = Store Category Sub Name (Countries)
 - Price = Current Display Price (Canadian \$)
 - Quantity = Total Weekly Selling Unit (Bottles)
- Data Visualization
 - Bar, Box, Point, Line, Histogram, Density
- Data Analysis
 - Regression: Price Elasticity
 - Classification Econ Comp in the Cloud

future esn Ways process first gleave nets addresses package parallel concerned statistics allows google second term theory issues distribution de examples per access inference hadoop systems working e separate described combined models describe include examples measure file ase transactions possible conditional scientists friedman queries machines general visualization interested gsmallh databases subsets Sens information residuals valid valid valid op op dim dim commonly large mentation absolute gclassififi available regression california decision journal validation datasets learning fif exploratory conventional computer typically statisticians linear economics handling dboodb spreadsheet iid structured according opensource relationships value additional nonlinear across language manipu gtallh dimensions processing specialists appendix classifi involved Cossvalidation better cloud summarization environment standard application K computing tricks parameter performance programming flfl extracted economic relative techniques gnot variable Known technique mapreduce tools mean econometricians econometric datacleaning common economists predictors applications prediction methods model billion ata larger testing measures econometrics Varian vector means various insample observations third gmappedh gfeaturesh estimation outofsample perspectives sql sample manipulation tasks variables mining predictions primarily machine gcarth gpredictorsh generally to associated particularly demanding given choose collaborations computers cost increasing gfath hal economist classififi using independent necessary want function introduce often appropriate USEd gpartition analyze analysis reasonably analyzing random outcome consider example query table depicted figure different provided database usually amazon option thousands number among big related training east tuning terms become dremel divide require shown high values start

