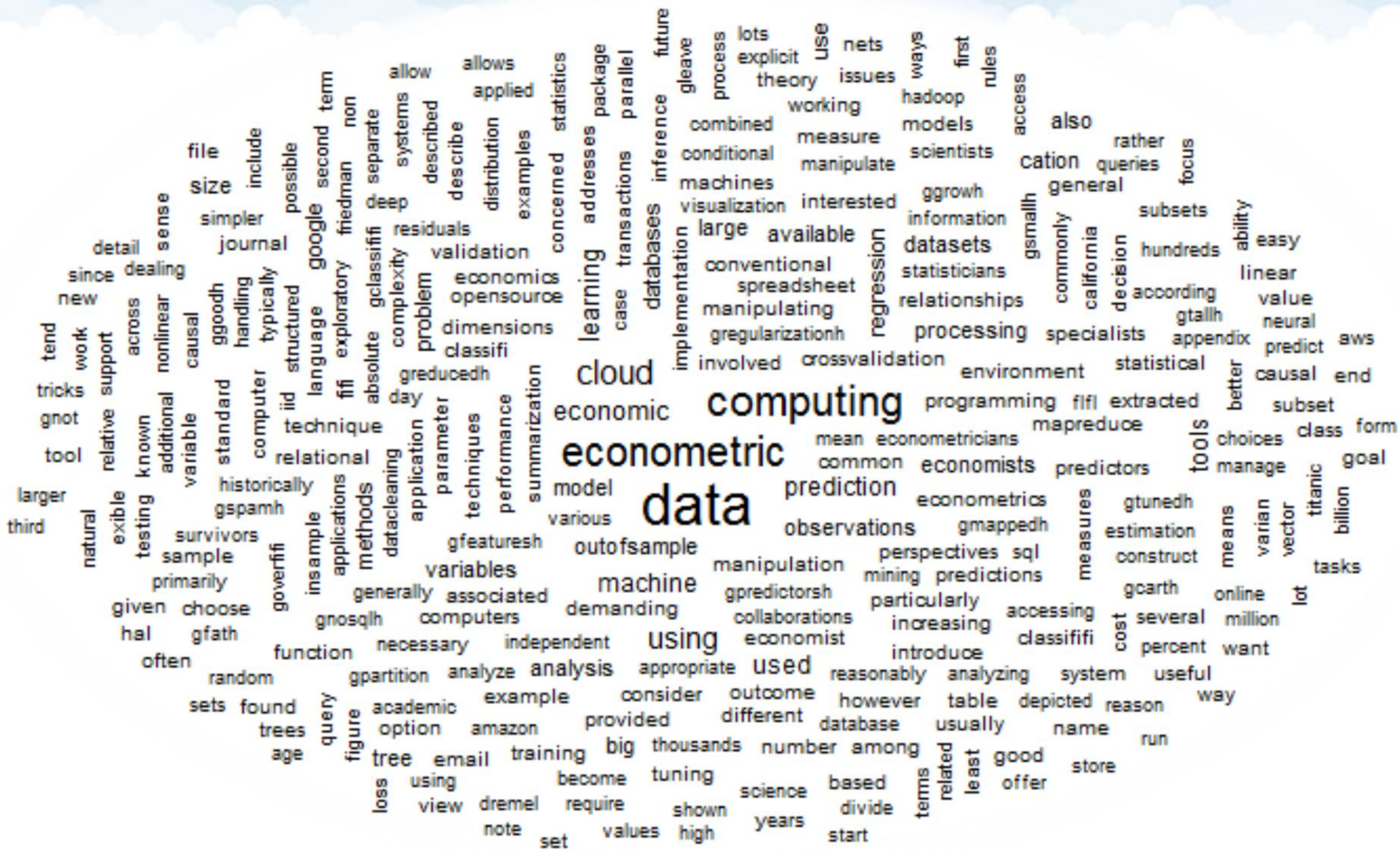
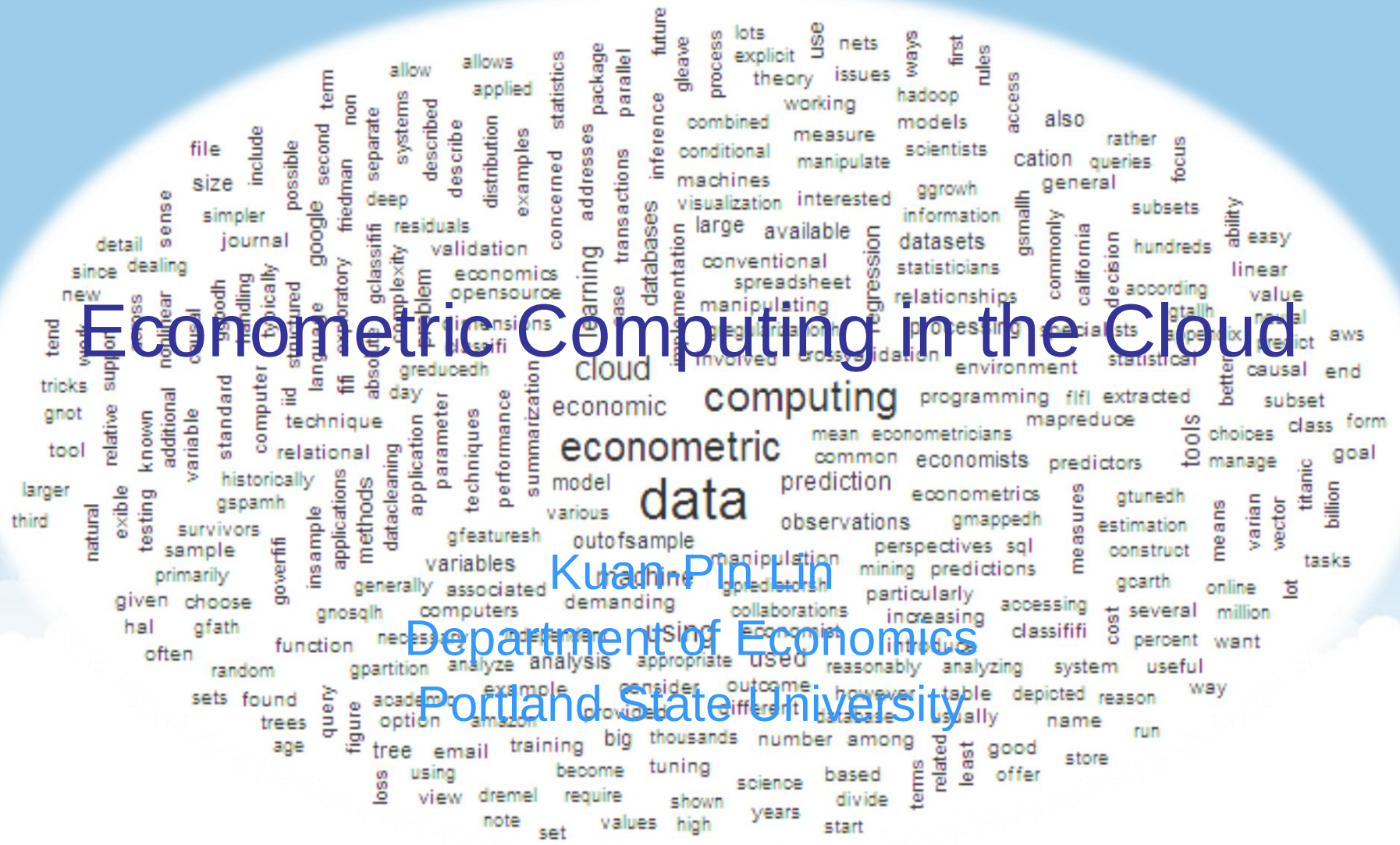

data
computing
econometric
cloud
economic
using
machine
prediction
learning
observations
economists
econometrics
used
analysis
model
outoffsample
variables
economist
manipulation
processing
datasets
regression
large
available
tools
also
cation
general
models
measure
interested
machines
combined
conditional
visualization
conventional
spreadsheet
manipulating
gregularizationh
involved
crossvalidation
environment
programming
mapreduce
econometricians
mean
common
predictors
statistical
extracted
flfl
relationships
statisticians
information
ggrowh
scientists
manipulate
working
theory
explicit
lots
nets
issues
hadoop
use
ways
first
rules
process
gleave
future
parallel
package
statistics
allows
allow
applied
inference
databases
transactions
addresses
concerned
examples
distribution
describe
described
systems
separate
non
term
second
possible
include
file
size
simpler
journal
detail
since
dealing
new
sense
google
friedman
deep
residuals
validation
economics
opensource
dimensions
classifi
greducedh
day
technique
relational
historically
gspamh
survivors
sample
primarily
given
choose
hal
gfath
often
random
sets
found
trees
age
query
figure
tree
loss
using
view
email
note
set
dremel
values
high
require
become
tuning
shown
years
science
based
divide
start
terms
related
least
good
offer
store
run
name
reason
way
useful
system
depicted
usually
table
however
database
number
among
thousands
big
training
amazon
option
academic
example
provided
consider
different
outcome
appropriate
analyze
gpartition
function
necessary
independent
computers
gnosqlh
generally
associated
demanding
collaborations
gpredictorsh
particularly
increasing
introduce
reasonably
analyzing
classififi
accessing
cost
several
percent
want
million
online
lot
gcarth
construct
estimation
gtunedh
measures
means
varian
vector
titanic
billion
tasks
goal
manage
choices
class
form
subset
causal
end
better
predict
aws
neural
appendix
gtallh
according
value
linear
decision
california
commonly
gsmallh
hundreds
subsets
ability
easy
focus
queries
rather
access
specialists
mining
predictions
perspectives
sql
gmappedh
various
gfeaturesh
datacleaning
application
parameter
techniques
performance
summarization
methods
applications
insample
goverfifi
testing
exible
natural
third
larger
tool
gnot
tricks
tend
work
support
relative
known
additional
variable
standard
computer
iid
fifi
absolute
language
structured
typically
handling
ggoodh
causal
nonlinear
across
exploratory
gclassififi
complexity
problem
case
implementation

Econometric Computing in the Cloud



Kuan-Pin Lin

Department of Economics

Portland State University

Econometric Computing

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

Econometric Computing

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

Using R

- Why R (and RStudio)?
  - Powerful, Popular, and Portable
  - Free (Nothing better than that!)
  - But, with steep learning curve!
- Using R Packages
  - lm(), 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, IaaS
    - IaaS: 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

Econometric Computing in the Cloud

- A Low Cost Option
  - Amazon Web Services (AWS: Free Tier)
    - EC2 with 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)
  - Data Source: American Association of Wine Economists

An Example

- Wine Sales in Vancouver BC
  - 37228 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

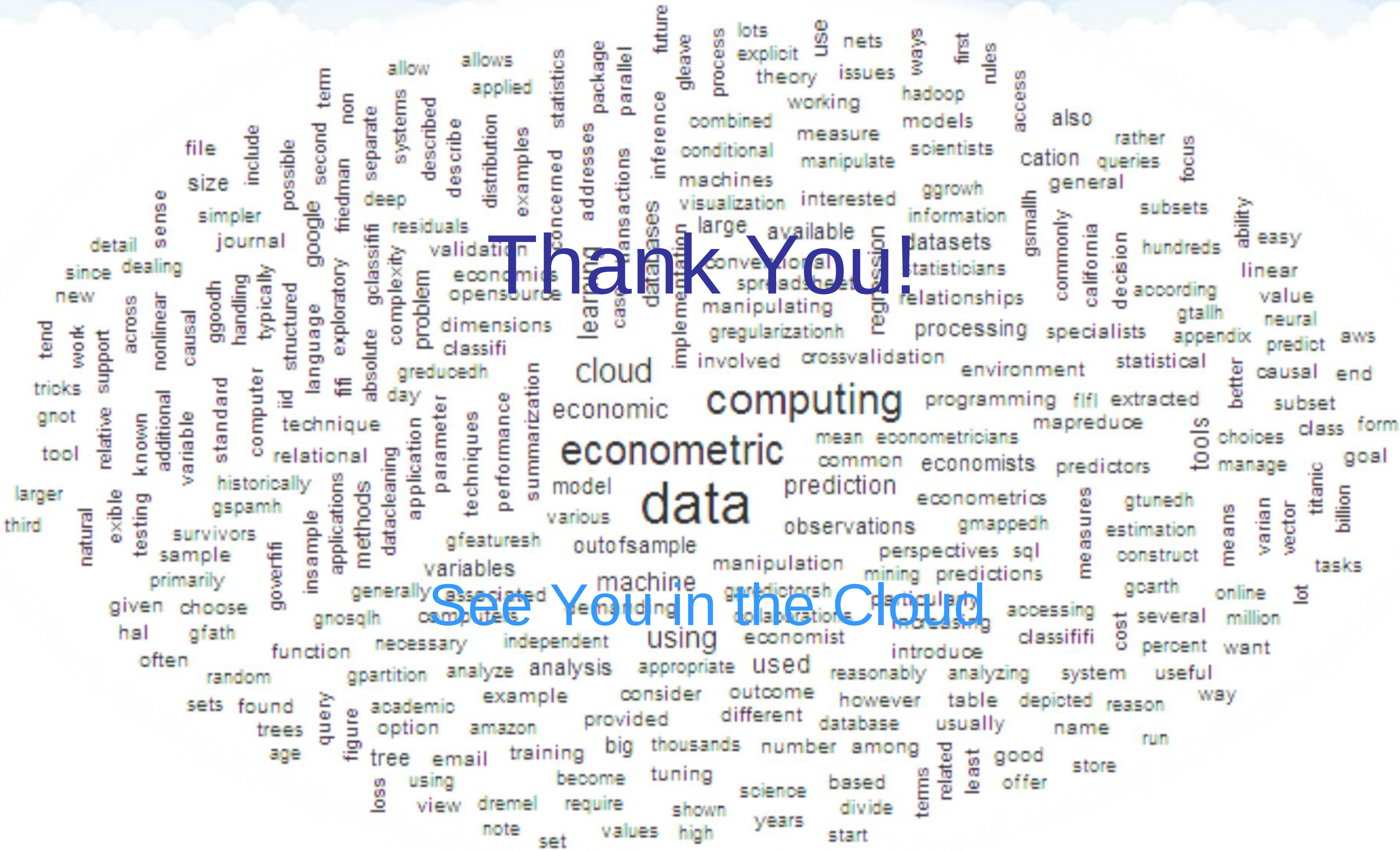
An Example

- 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


data
computing
econometric
cloud
learning
economic
machine
using
prediction
model
analysis
used
observations
economists
econometrics
outofsample
manipulation
processing
datasets
regression
databases
transactions
implementation
large
available
conventional
spreadsheet
statisticians
relationships
environment
programming
mapreduce
predictors
tools
statistical
specialists
economics
opensource
dimensions
classifi
validation
technique
relational
variables
machines
visualization
interested
information
measure
models
scientists
hadoop
statistics
package
parallel
future
process
inference
addresses
concerned
examples
distribution
describe
described
systems
separate
friedman
google
language
exploratory
structured
typically
handling
standard
computer
variable
additional
nonlinear
causal
support
sample
survivors
primarily
choose
given
often
random
function
necessary
independent
computers
associated
generally
demanding
appropriate
consider
outcome
different
reasonably
analyzing
system
useful
however
table
depicted
reason
database
usually
name
number
among
thousands
provided
example
academic
option
amazon
training
email
tree
big
tuning
science
based
divide
years
start
shown
high
values
require
become
dremel
note
view
set
store
offer
good
least
related
terms
run
way

[Let’s Go!]

Thank You!



See You in the Cloud