



Dyalog North America – 29-30 September 2025

The Data Science Journey

Josh David

Why the need

- Heavily used in rising fields
 - Data Science
 - Machine Learning
- Shift in Company Cultures
 - Data driven
- Opportunity for APL

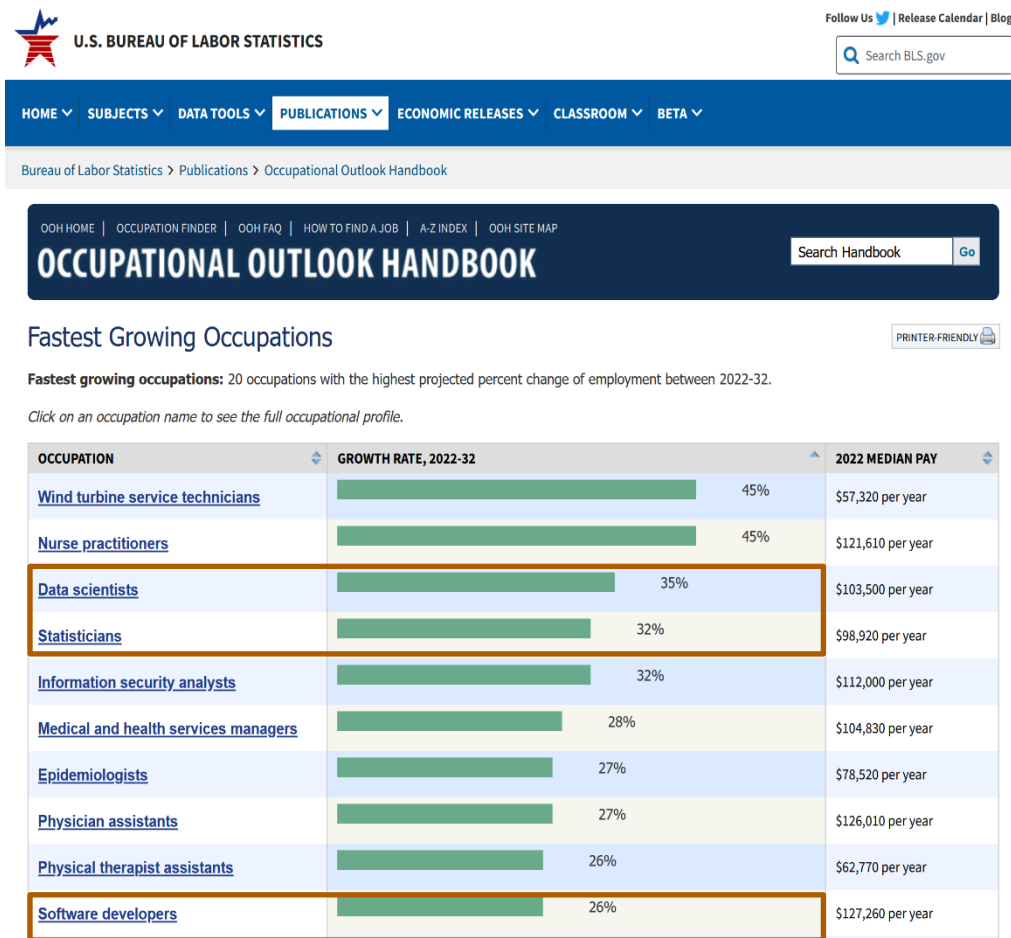


Figure 1



Fastest Growing Occupations

PRINTER-FRIENDLY

Fastest growing occupations: 20 occupations with the highest projected percent change of employment between 2022-32.

Click on an occupation name to see the full occupational profile.

OCCUPATION	GROWTH RATE, 2022-32	2022 MEDIAN PAY
Wind turbine service technicians	45%	\$57,320 per year
Nurse practitioners	45%	\$121,610 per year
Data scientists	35%	\$103,500 per year
Statisticians	32%	\$98,920 per year
Information security analysts	32%	\$112,000 per year
Medical and health services managers	28%	\$104,830 per year
Epidemiologists	27%	\$78,520 per year
Physician assistants	27%	\$126,010 per year
Physical therapist assistants	26%	\$62,770 per year
Software developers	26%	\$127,260 per year

Fastest Growing Occupations

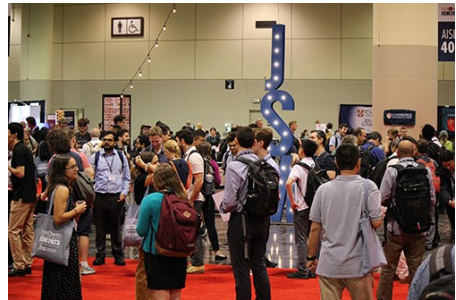
PRINTER-FRIENDLY

Fastest growing occupations: 20 occupations with the highest projected percent change of employment between 2024-34.

Click on an occupation name to see the full occupational profile.

OCCUPATION	GROWTH RATE, 2024-34	2024 MEDIAN PAY
Wind turbine service technicians	50%	\$62,580 per year
Solar photovoltaic installers	42%	\$51,860 per year
Nurse practitioners	40%	\$129,210 per year
Data scientists	34%	\$112,590 per year
Information security analysts	29%	\$124,910 per year
Medical and health services managers	23%	\$117,960 per year
Physical therapist assistants	22%	\$65,510 per year
Actuaries	22%	\$125,770 per year
Operations research analysts	21%	\$91,290 per year
Physician assistants	20%	\$133,260 per year
Psychiatric technicians	20%	\$42,590 per year

3



DYNA



JSM

STATISTICS

Statistician weighs 155 lbs with probability of 25 lbs. What is the probability that a randomly selected statistician will weigh over 170 lbs?

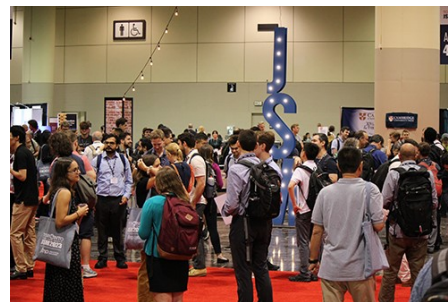
TS
normal probability > 170

X
DIST(170,155,25,TRUE)

Q
55,25,lower.tail=FALSE)

ANOTHER CALC PROGRAM?
ifies statistics by using one distribution, unlike R or Excel. It has four core functions: **any, distribution, relational,** and introduces operators that allow users to perform tasks like simulations, running hypothesis tests or regression.

DVALOC



General Purpose

Quantum Computing Research



Researchers at the National Center for Superconducting Applications (NCSA) have developed a quantum circuit that can be used to solve a wide range of problems. The circuit is designed to solve a problem that is known as the "NCSA problem". The circuit is designed to solve a problem that is known as the "NCSA problem".

Mine Processing Model



Conventional models are based on the use of a single data source. The model is designed to solve a problem that is known as the "NCSA problem". The model is designed to solve a problem that is known as the "NCSA problem".

Taming Statistics



Statistics is a field of study that deals with the collection, analysis, and interpretation of data. The field is divided into two main branches: descriptive statistics and inferential statistics. Descriptive statistics is the study of the characteristics of a group of data, while inferential statistics is the study of the characteristics of a single data point.

Numerical Modelling of Ocean Currents



Researchers at the National Center for Superconducting Applications (NCSA) have developed a numerical model of ocean currents. The model is designed to solve a problem that is known as the "NCSA problem". The model is designed to solve a problem that is known as the "NCSA problem".

DNA Analysis

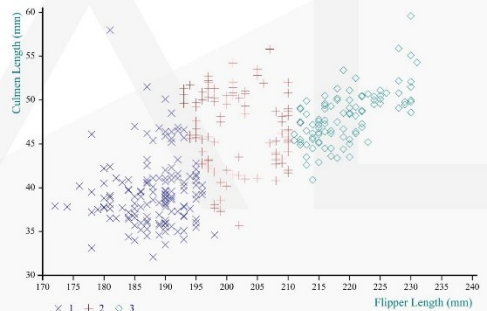


Researchers at the National Center for Superconducting Applications (NCSA) have developed a DNA analysis. The analysis is designed to solve a problem that is known as the "NCSA problem". The analysis is designed to solve a problem that is known as the "NCSA problem".

Dyalog's data manipulation and rapid development capabilities allow users to respond quickly to user requests and changing business requirements.

Symbolic Notation

ab	$a \times b$	$\min(x_0, \dots, x_n)$	L/x
$\frac{x}{y}$	$x \div y$	$\max(x_0, \dots, x_n)$	$\lceil x \rceil$
$\sum_{i=1}^n i$	$+ / 4 \times 16$	$\lceil x \rceil$	$\lceil x \rceil$
$\prod_{i=1}^6 i$	$\times / 4 \times 16$	$\lfloor x \rfloor$	$\lfloor x \rfloor$
e^x	$\star x$	$\sum_{i=1}^N x_i$	$+ / \div \neq$
x^y	$\star \star y$	$\frac{1}{N} \sum_{i=1}^N x_i$	$+ / \div \div$
$\log_e(x)$	$\otimes x$	$\sum_{i=1}^N a_i b_i$	$+ \cdot \times$
$\log_{10}(x)$	$10 \otimes x$	$\sum_{i=1}^N w_i x_i$	$+ \cdot \times \div \div \div$
$n!$	$! n$	$\frac{P(A B_i)P(B_i)}{\sum P(A B_j)P(B_j)}$	$\times \div + \cdot \times$
$\binom{n}{k}$	$k ! n$		
$\sqrt{a}$	$a \star \div n$		
$f g x$	$f g x$		



K Means Clustering of Palmer Archipelago Penguins

```

m[kg]=x[+/#]B[a-g]=a[k]1=a.(-x2[2]u)*a[m]1=37#m=f,l,z,l
c=k[m]1=37#m=f,l,c,l
Converge={
  d=a.(-x2[2]u)
  g=a[k]1=d
  kg=(+/#)B[a]
}
m Converge=a c
  
```

A Initial centroid is random
 A Distance to centroids
 A Group by nearest centroid
 A New centroid is mean of corresponding group

A Tool of Thought

Concise

Simple syntax and symbols used to represent common operations on data allow users to write and adapt short and elegant expressions for all kinds of purposes. Commonly used patterns and expressions are called "idioms".

Mean average

$2.8 \quad (+/#)3 \ 1 \ 4 \ 1 \ 5$

Windowed moving average

$2.2 \ 5 \ 2.5 \ 3$

Weighted average

2.666666667

Integrated

Dyalog-based software can be deployed as scripts, graphical desktop applications, web applications and services hosted in the cloud, or integrated as part of an existing technology stack.

- Read and write data in CSV, JSON and XML formats
- Read, write and manage SQL databases
- Read and write arbitrary data types
- Interface directly with Microsoft Excel and other Microsoft Office products for automation
- Provide and consume web services
- Create compiled libraries (.dll, .so, .dylib), and use those written in C or other languages
- Use and create .NET assemblies to interoperate with C# and other .NET languages
- Talk to live-running Python and R systems

Performant

Often completely branchless, APL expressions present a high degree of mechanical sympathy. Ideally suited to SIMD processing, APL can offer high programmer efficiency, as well as all-out execution speed by leveraging modern processors with dedicated vector-oriented instructions.

Which are vowels?

$'aesthetic'c'aelou'$

$1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 1 \ 0$

Only leading vowels

$\wedge'/'aesthetic'c'aelou'$

$1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0$

Remove leading hyphens

$(w^2-\wedge'(-'u)')---sample \ text---$

sample text---

Remove trailing hyphens

$(w^2-\wedge'(-'u)')---sample \ text---$

$---sample \ text$

Expressive

Common patterns in APL can be applied in many use cases. Conversely, there are often many ways to tackle the same problem. The small code size lowers the cost of trying multiple approaches to find what works best for your application.

A windowed plus-reduction gives the sum of each set of three consecutive numbers.

$8 \ 6 \ 10 \ 15$

A windowed catenate-reduction returns the groupings as a list of lists

$3, / 2 \ 1 \ 4 \ 1 \ 5 \ 9$

$3 \ 1 \ 4 \ 1 \ 4 \ 1 \ 4 \ 1 \ 5 \ 1 \ 5 \ 9$

A windowed less-than-reduction keeps the first 1 in each consecutive group of 1s in a Boolean array.

$1, 2, 2 / 'a' mark the start'$

$1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0$

Simple

Primitive functions and operators are represented by symbols that mnemonically suggest their meanings.

$'a' \text{ 'applies'}$

$6 \quad 'reverse'$

$asrev \quad 'rotate'$

$atrot \quad '3, 1, 1'$

$4 \quad '3, 1, 1'$

$5 \quad '3, 1, 1'$

Getting Started

Dyalog

Download the latest version of Dyalog:

www.dyalog.com/downloads

TryAPL

TryAPL includes basic APL tutorials, glyph input, and an interactive workspace for experimenting with the language.

www.tryAPL.com

TamStat

TamStat can be downloaded from:

tamstat.dyalog.com

The APL Challenge

Learn APL and compete for one of three \$100 prizes.

challenge.dyalog.com

The APL Forge

Got a problem to solve or a library to share? Use Dyalog APL to build something great – and win \$2,500 + an expenses-paid trip to present at a Dyalog user meeting.

forge.dyalog.com

sales@dyalog.com

DYALOG

www.dyalog.com

DYNA

Symbolic Notation

ab	$a \times b$	$\min(x_0 \dots x_n)$	$\lfloor / x$
$\frac{x}{y}$	$x \div y$	$\max(x_0 \dots x_n)$	$\lceil / x$
$\sum_{n=1}^6 4n$	$+ / 4 \times 16$	$\lceil x \rceil$	$\lceil x$
$\prod_{i=1}^6 4i$	$\times / 4 \times 16$	$\lfloor x \rfloor$	$\lfloor x$
e^x	$* x$	$\frac{\sum_i^N x_i}{N}$	$+ / \div \neq$
x^y	$x * y$	$\frac{1}{k} \sum_{i=n-k+1}^n x_i$	$+ / \div \neg$
$\log_e(x)$	$\otimes x$	$\sum a_i b_i$	$+ . \times$
$\log_{10}(x)$	$10 \otimes x$	$\frac{\sum w_i x_i}{\sum w_i}$	$+ . \times \div + / \ddot{\circ} \neg$
$n!$	$! n$	$\frac{P(A B_i)P(B_i)}{\sum P(A B_i)P(B_i)}$	$\times \div + . \times$
$\binom{n}{k}$	$k ! n$		
$\sqrt[n]{a}$	$a * \div n$		
$f g x$	$f \ g \ x$		

K Means Clustering

$m\{\phi g \vdash \circ (+ \neq \div \neq) \exists \alpha \vdash g \leftarrow \supset \circ \Delta \ddot{\circ} 1 \vdash \alpha + . (* 2 \ddot{\sim}) \omega\} * \equiv \phi m \square \ddot{\sim} c 3 ? \neq m \leftarrow f l , \overline{\sim} c l$

K Means Clustering

$c \leftarrow \text{random}(m)$ Initial centroid is random

Converge $\leftarrow \{$

$d \leftarrow \text{distance}(x, c)$ Distance to centroids

$g \leftarrow \text{group}(x, d)$ Group by nearest centroid

$c_g \leftarrow \text{mean}(g)$ New centroid is mean of

$\}$ corresponding group

return Converge

What is Data Science?

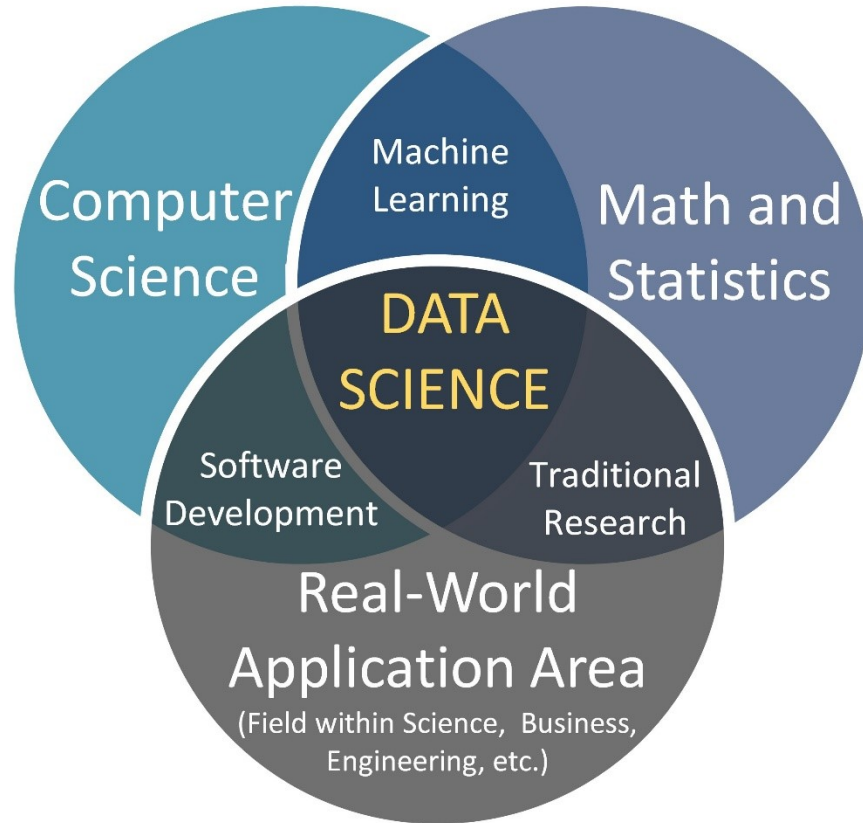
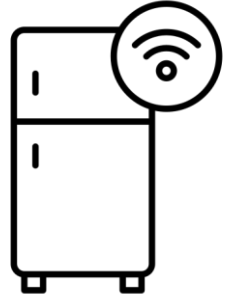
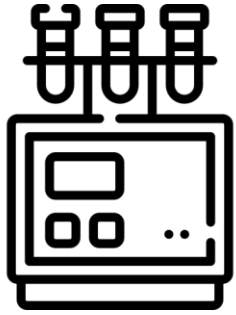


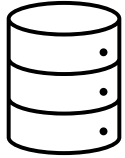
Figure 3

Data Science is the Collection of Data



Data Science is the Structuring of Data

Databases



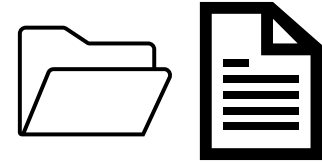
)load sqapl

APIs



]load HttpCommand

Files

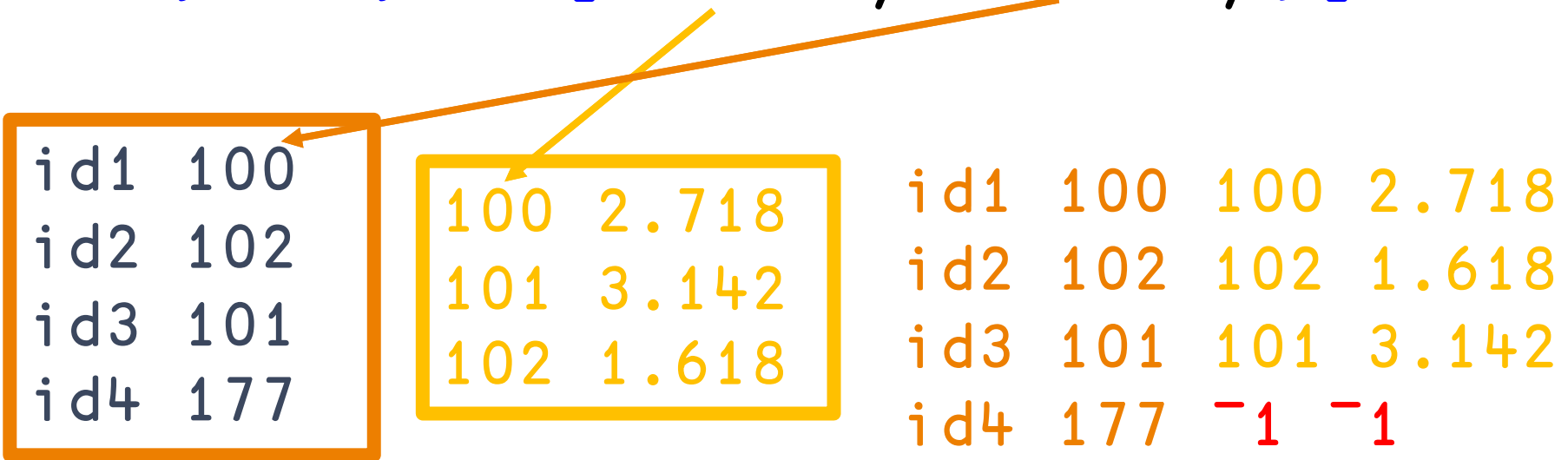


☐ CSV
☐ JSON
☐ XML

workspace

Data Science is the Structuring of Data

`t1, (t2, -1)[t2Pkey, t1Fkey:]`



id1	100
id2	102
id3	101
id4	177

100	2.718
101	3.142
102	1.618

id1	100	100	2.718
id2	102	102	1.618
id3	101	101	3.142
id4	177	-1	-1

Data Science is the **Cleansing** of Data

$x \sim \square \text{NULL}$ $\#$ X without nulls

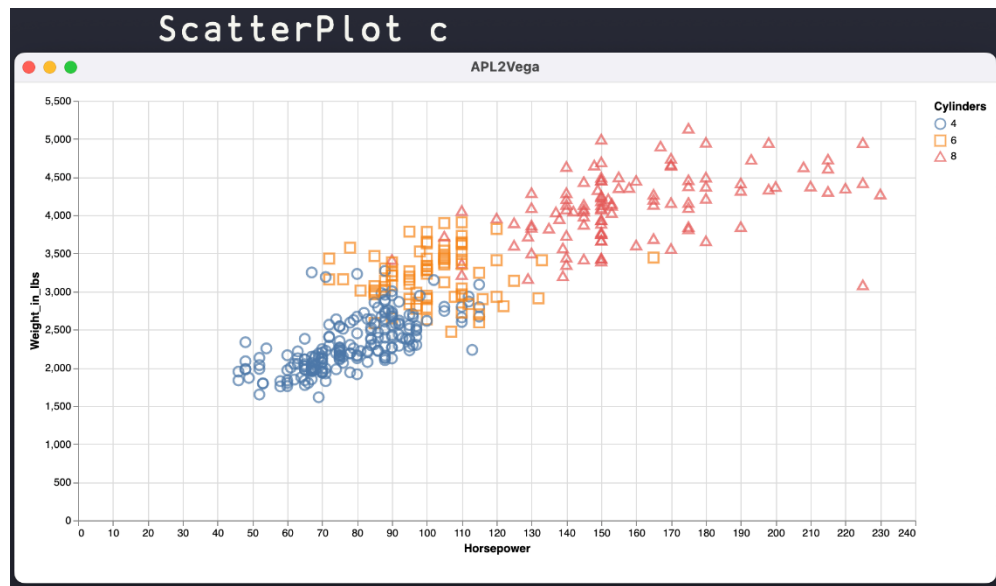
$x / \sim (x > 0) \wedge (x \leq 100)$ $\#$ Keep values between (0,100]

$u x$ $\#$ Remove duplicates

$tlb \leftarrow \{ (v \setminus ' ' \neq \omega) / \omega \}$ $\#$ Trim leading blanks

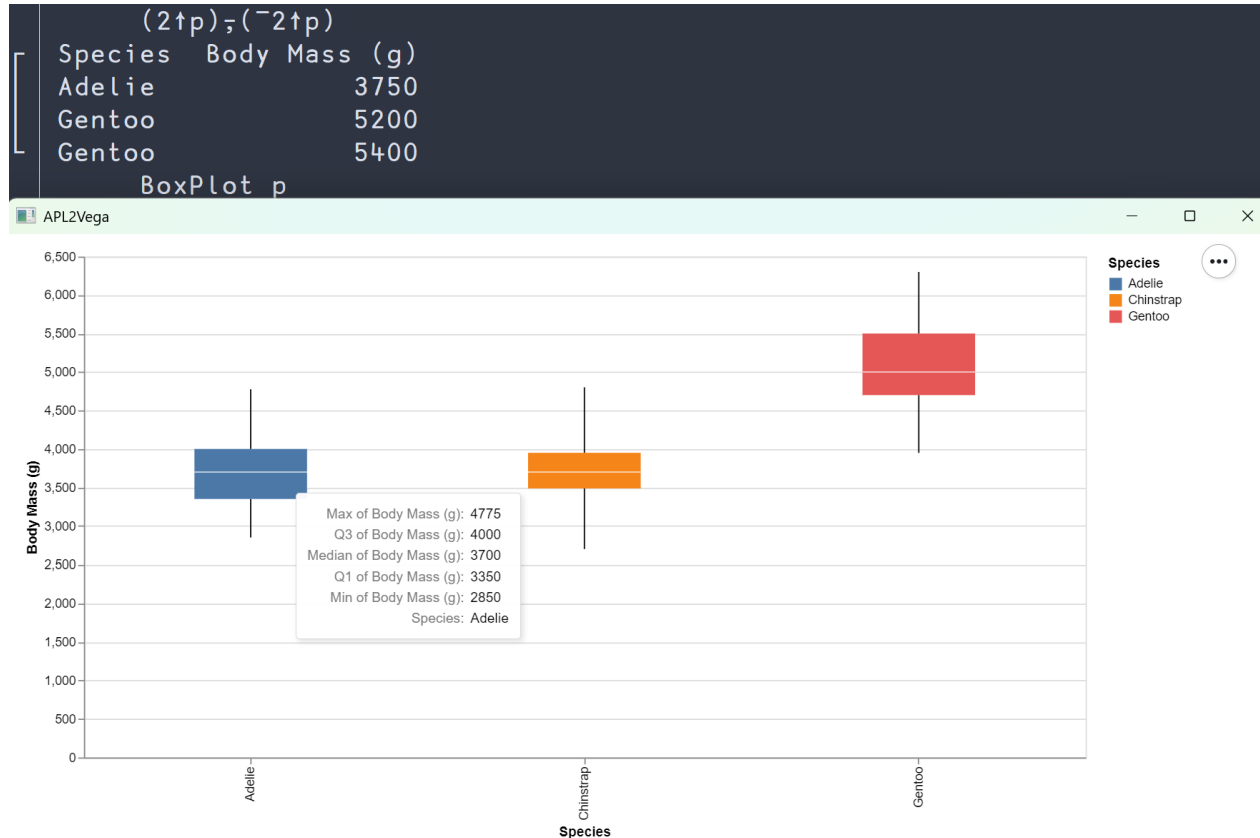
$tlb \circ \phi * 2$ $\#$ Trim trailing and leading blanks

Data Science is the Visualization of Data

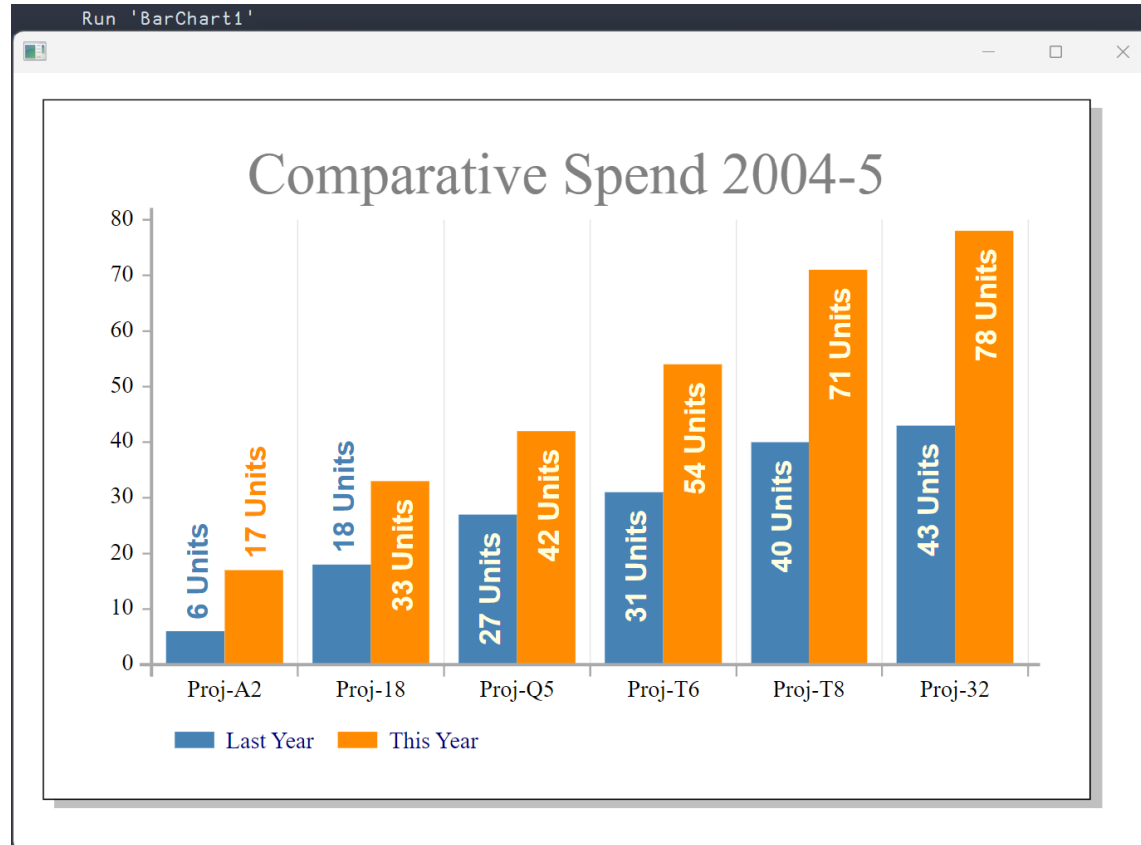


<https://vega.github.io/vega-lite/examples>

Data Science is the Visualization of Data



Data Science is the Visualization of Data



Data Science is the Summarization of Data

$+/$	A sum
$\times/$	A product
$\lceil /$	A max
$\lfloor /$	A min
$(\lceil / - \lfloor /)$	A range
$(+ / \div \neq)$	A mean

Data Science is

- ◆ Data Science is an exploratory discipline, combining fields of Mathematics, Statistics & Computing
- ◆ Sounds perfect for APL?

History – GRAFSTAT (1984)

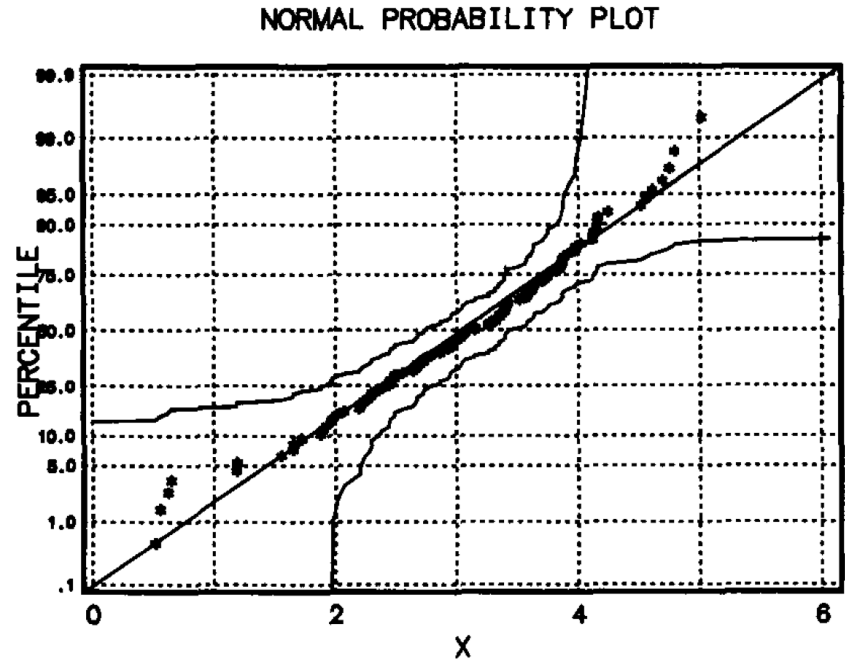
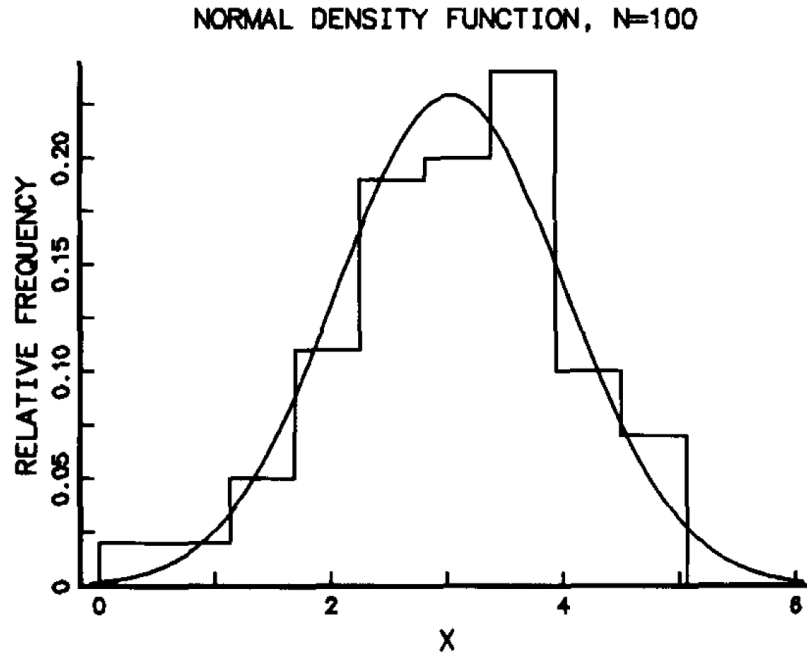
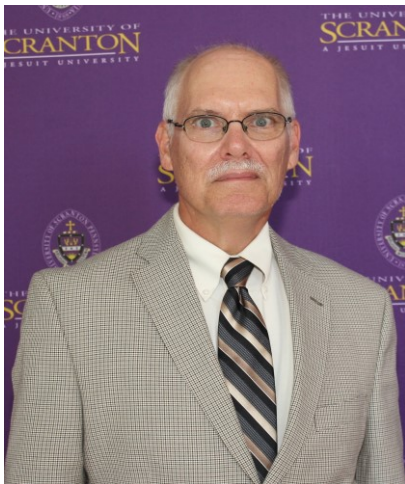


Figure 4

Open Source Libraries



KokoStats



TamStat



APLearn

[About the APL Forge](#)[Previous Winners](#)[Submissions](#)[Ideas](#)[FAQs](#)[Contact Us](#)[SUBMIT YOUR PROJECT](#)

Shaping tomorrow with APL

Forge your ideas into prize-winning APL tools

Do you have a problem and are looking for the right language in which to solve it? Do you have an idea for a library that you think other APL users might benefit from? Do you have an APL-based application that you want to share? The APL Forge is where we reward you for using Dyalog APL to solve problems and develop libraries, applications, and tools.

[MORE DETAILS](#)

KokoStats

github.com/JoshDavid/KokoStats

Anova, AnovaPool	PrincipleComponents	DFT
RegressMultipleLinear	Statistics	FFT
RegressPolynomial	CorrelMatrix	IDFT
RegressForsythe	AutoCorr	IFFT
RegressChebyshev	CrossCor	TukeyWindow
RegressFourier	CrossTabs	LeadLag

TamStat

A full-stack application with a functional DSL for statistical operations

TamStat

github.com/steveman7/TamStatCore
tamstat.dyalog.com

Probability that a person weighs over 150 lbs given a population mean of 120 and a standard deviation of 25 lbs?



`pnorm(150,120,25,lower.tail=FALSE)`



`=1-NORM.DIST(150,120,25,TRUE)`



`120 25 normal probability > 150`

TamStat

What is the probability of at least 4 heads in 5 tosses?

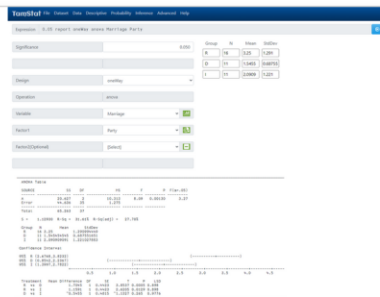
 `pbinom(4-1,5,0.5,lower.tail=FALSE)`

 `=1-BINOM.DIST(4-1,5,0.5,TRUE)`

 5 0.5 binomial probability ≥ 4

0.1875

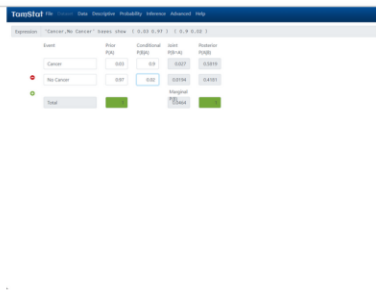




Anova Wizard

Marriage only between man and woman?

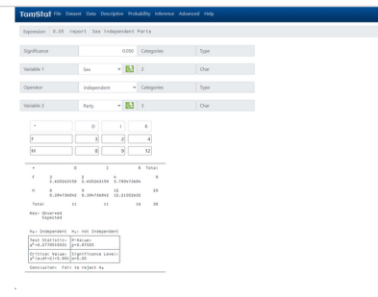
View



Bayes Wizard

Probability of having cancer

View



ChiSquare Wizard

Party affiliation & sex

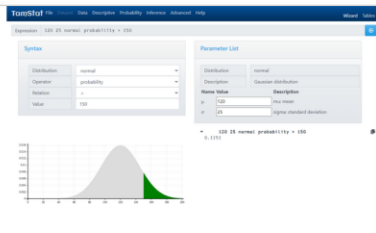
View



Confidence Interval Wizard

Estimate mean height

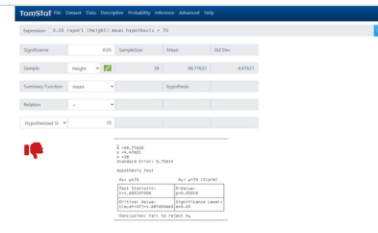
View



Distribution Wizard

Probability of weight > 150 lbs.

View



Hypothesis Wizard

Average adult height < 70 inches?

View

TamStat - Demo

Issues with ML Libraries

- High level libraries that abstract away the central logic of a program
- Low-level code that obfuscate the core definition of an algorithm



SciPy

APLearn - Demo

Sci-kit learn, written in APL

Baseline Model Performance

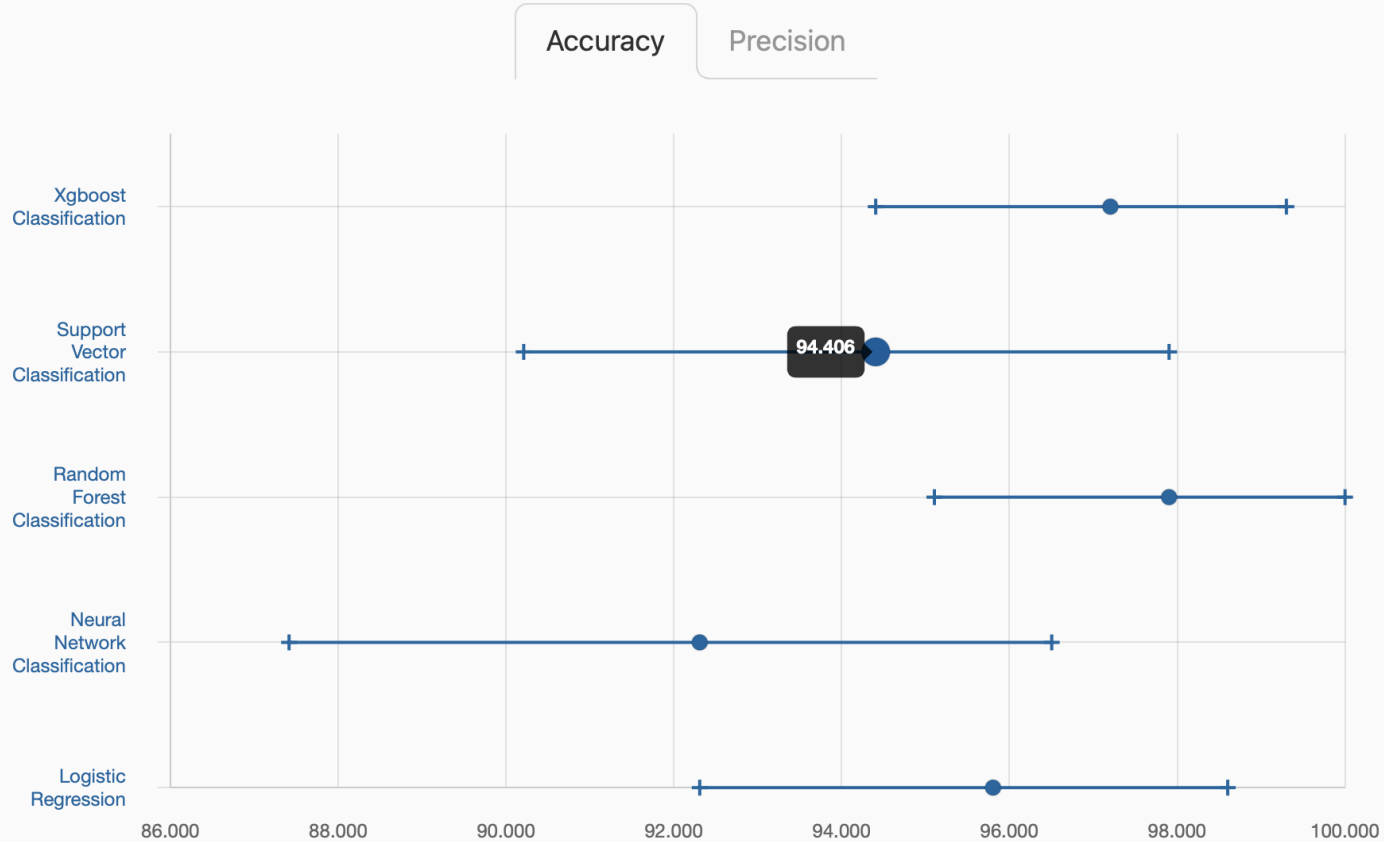


Figure 5

Baseline Model Performance

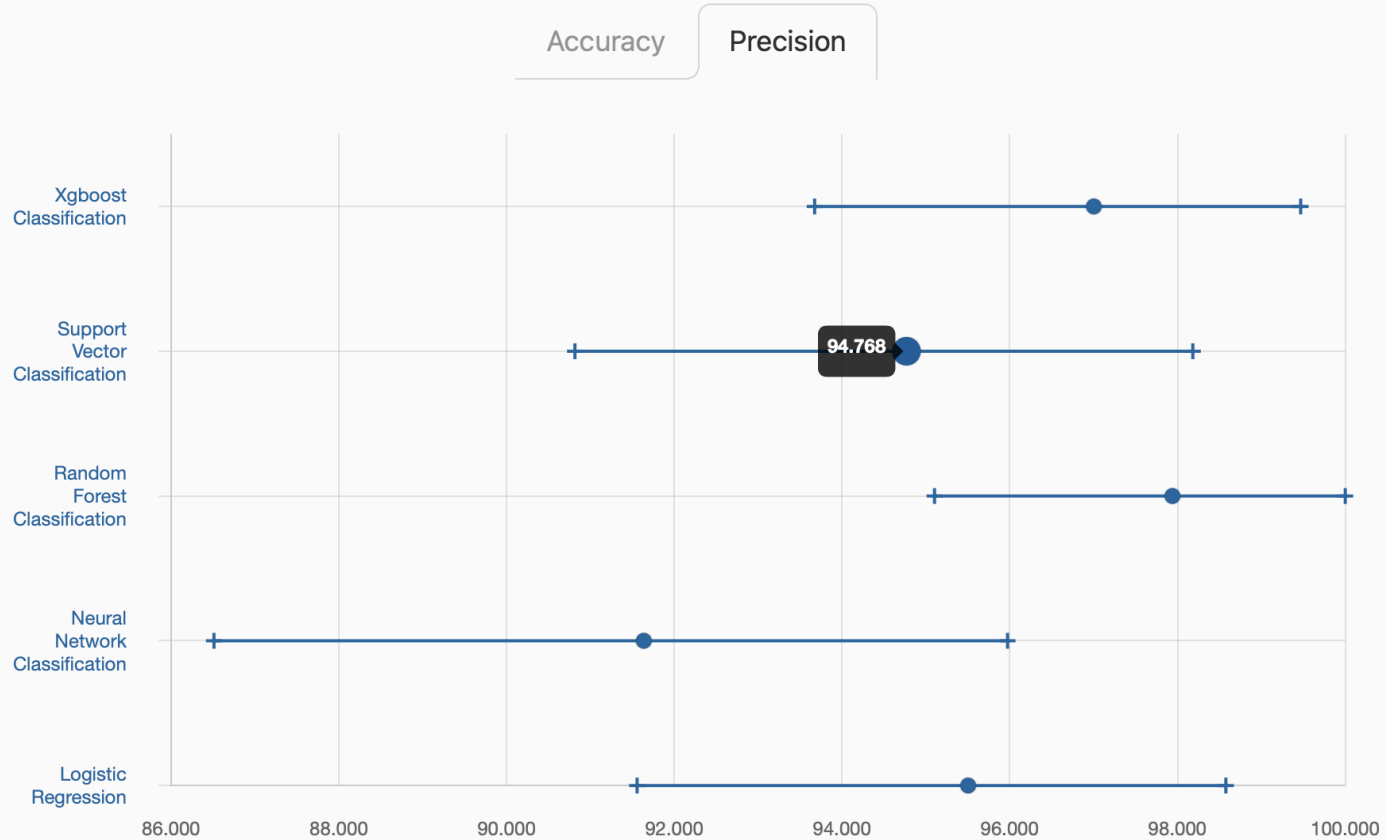


Figure 6

Key (v16+)

$\{(3 \uparrow \omega) \dot{-} (\bar{3} \uparrow \omega)\}$ borough, $\dot{-}$ status

Queens	Alive
Queens	Alive
Brooklyn	Alive
Staten Island	Alive
Bronx	Alive
Queens	Alive

Key 目 (v16+)

{ α , $\neq \omega$ } 目 borough status

Queens	250551	Alive	652173
Brooklyn	177293	Stump	17654
Manhattan	65423	Dead	13961
Staten Island	105318		
Bronx	85203		

Key 目 (v16+)

borough{c $\leftarrow$ ω ◇ α, c{α, 100×c÷ω}目ω}目status

Brooklyn	Alive	95.74207668
	Dead	1.872042325
	Stump	2.385880999
Manhattan	Alive	95.42057075
	Dead	2.754383015
	Stump	1.825046238

...

Stencil $\boxtimes$ (v16+)

max

$\{\Gamma \neq, \omega\} \boxtimes (2 \times 2 \rho^2) \vdash m$

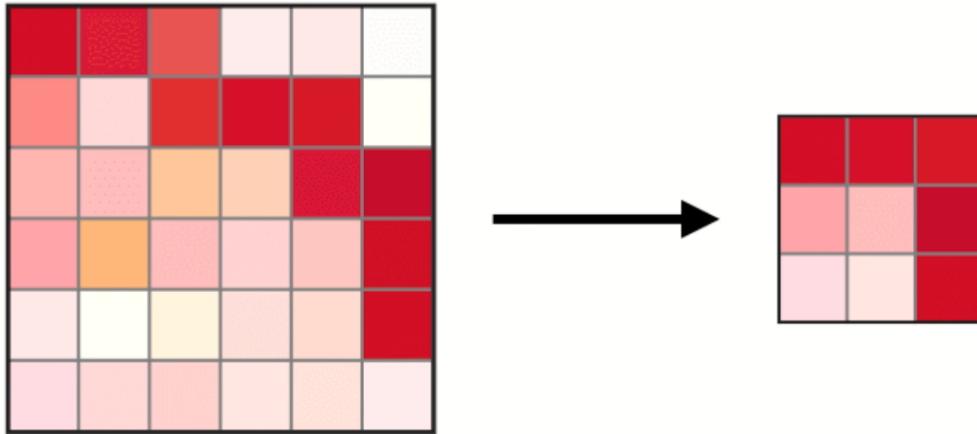


Figure 7

U-net CNN

We're building this...

(obviously!)

U-net CNN in APL

Exploring zero-framework, zero-library machine learning

... from scratch.

W←θ ∘ V←θ ∘ Z←θ ∘ LR←1e⁻⁹ ∘ MO←0.99

```
FWD←{Z←+(#W)ρ<θ
CV←{Z[α]←cω
z←(,[2+13]3 3⊗α>Z)+.×,[13]α>W
0[z-z[α]←cZ[α],←z}
CC←{p←2÷(ρα>Z)-ρω
ω,∘([p]+(-[p])+(α>Z))
MX←{f←[2],[2 3](2 2p2)⊗>Z[α]←cω}
UP←{Z[α]←cω
s←(2×1+ρω),-1↑ρα>W
sp0 2 1 3 4qω+.×α>W}
C1←{Z[α]←cω
1E-8+z÷[12]+/z×z-[12][z+ω+.×α>W}
LA←{α>#Z:ω
down←(α+6)▽(α+2)MX(α+1)CV(α+0)CV ω
(α+2)CC(α+5)UP(α+4)CV(α+3)CV down}
2 C1 1 CV 0 CV 3 LA ωρ÷3↑1,∘ρω}

BCK←{Y←α ∘ YΔ←ω
Δ←{V[α]←cω+MO×(ρω)ρα>V
W[α]←c(α>W)-LR×α>V}
ΔCV←{w←,[13]⊗ϕ[1]0 1 3 2qα>W ∘ x←α>Z
Δz←ω×0<1>α>Z
ΔZ←-2e-2ϕ[1](4+2↑ρΔz)↑Δz
←α Δ 3 0 1 2q(q,[12]Δz)+.×,[12]3 3⊗x
w+.×,∘[2+13]3 3⊗ΔZ}
ΔCC←{x←α>Z ∘ Δz←ω ∘ d←-[2÷2↑(ρx)-ρΔz
(α>d)⊗(1>d)ϕ[1](ρx)↑Δz}
ΔMX←{x←α>Z ∘ Δz←ω
y×x←y-(ρx)↑2÷2/[1]Δz}
ΔUP←{w←α>W ∘ x←α>Z ∘ Δz←ω
←α Δ(q,[12]x)+.×,[12]cz+(2 2p2)⊗Δz
([2+13]cz)+.×q;w}
ΔC1←{w←α>W ∘ x←α>Z ∘ Δz←ω
←α Δ(q,[12]x)+.×,[12]Δz
Δz+.×q;w}
ΔLA←{α>#Z:ω ∘ in←ω↑[2]∘-2÷2÷ϕρω
d←(α+6)▽(α+3)ΔCV(α+4)ΔCV(α+5)ΔUP in
(α+0)ΔCV(α+1)ΔCV(ω ΔCC÷α+2)+(α+2)ΔMX d
}
3 ΔLA 0 ΔCV 1 ΔCV 2 ΔC1 YΔ-(~Y),[1.5]Y}

E←-+÷,⊙(α×ω[;1])+(~α)×ω[;0]}
RUN←{Y←[0.5+nm↑ω÷2÷2(ρω)-nm+2↑ρYΔ←FWD α
Y YΔ(Y E YΔ)-Y BCK YΔ}
```

Figure 8

Conclusion

- ◆ APL's notation is ripe for tasks in Data Science and Machine Learning
- ◆ Data Science and Machine Learning give a new perspective and influence core language design
- ◆ Work on co-dfns & GPU is critical

References

- Figure 1

- <https://www.bls.gov/ooh/fastest-growing.htm>

- Figure 2

- <https://ww2.amstat.org/meetings/jsm/2025/spotlight.cfm>

- Figure 3

- <https://artsci.usu.edu/math-stats/datascience/>

References

- Figure 4
 - <https://dl.acm.org/doi/pdf/10.1145/800058.801082>
- Figure 5,6
 - <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>
- Figure 7
 - <https://dl.acm.org/doi/10.1145/3589246.3595371>
- Figure 8
 - <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks>

References

Icons

- Lab [Equipment icons created by Freepik - Flaticon](https://www.flaticon.com/free-icons/equipment "equipment icons")
- Survey [Survey icons created by Freepik - Flaticon](https://www.flaticon.com/free-icons/survey "survey icons")
- Watch [Smart watch icons created by Rlkas Dzihab - Flaticon](https://www.flaticon.com/free-icons/smart-watch "smart watch icons")
- Fridge [Smart fridge icons created by logisstudio - Flaticon](https://www.flaticon.com/free-icons/smart-fridge "smart fridge icons")
- Shopping [Shopping cart icons created by Freepik - Flaticon](https://www.flaticon.com/free-icons/shopping-cart "shopping cart icons")