PROBABLY

probably
TO PROGRAM

"To Write a series of coded software instructions to Control the operation of a computer or another machine"
CODING LOOKS LIKE...
### Diagram for the computation by the Engine of

<table>
<thead>
<tr>
<th>Number of Operation</th>
<th>Nature of Operation</th>
<th>Variables acted upon</th>
<th>Variables receiving results</th>
<th>Indication of change in the value on any Variable</th>
<th>Statement of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>×</td>
<td>$V_2 \times V_3$</td>
<td>$V_4, V_5, V_6$</td>
<td></td>
<td>$2^n$</td>
</tr>
<tr>
<td>2</td>
<td>−</td>
<td>$V_4 - V_1$</td>
<td>$2V_4$</td>
<td></td>
<td>$2^n - 1$</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>$V_5 + V_1$</td>
<td>$2V_5$</td>
<td></td>
<td>$2^n + 1$</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>$2V_5 + 2V_4$</td>
<td>$V_{11}$</td>
<td></td>
<td>$\frac{2^n - 1}{2}$</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>$V_{11} + V_2$</td>
<td>$2V_{11}$</td>
<td></td>
<td>$1 - \frac{2^n - 1}{2}$</td>
</tr>
<tr>
<td>6</td>
<td>−</td>
<td>$V_{13} - 2V_{11}$</td>
<td>$V_{13}$</td>
<td></td>
<td>$1 - \frac{2^n - 1}{2}$</td>
</tr>
<tr>
<td>7</td>
<td>−</td>
<td>$V_3 - V_1$</td>
<td>$V_{10}$</td>
<td></td>
<td>$n - 1 (= 3)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_1$</td>
<td>$V_2$</td>
<td>$V_3$</td>
<td>$V_4$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>n</td>
<td>2n</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**INSTRUCTIONS**

@lynnlangit Link
PROGRAM THIS!

- The number 731 would be represented by 7s, 3s, E.
- The number 804 would be represented by 8s, X, 4L.
- The number 107 followed by the number 51 would be represented by 1s, X, 7L, 5s, E.
QUIPU
TALKING KNOTS

a complex language, recorded in a 3D system, used for statistical AND narrative information
PROBABLY

Programming
"We are distracted..." - Erik Meijer
eleanor tweets about cats @noneuclideangrl · Aug 17

```
.length()
.length
.len()

LAWFUL GOOD

.getLength()
.count()
len()

NEUTRAL GOOD

.size()

CHAOTIC GOOD

.m_sizeGet()

LAWFUL NEUTRAL

TRUE NEUTRAL

sizeof()

LAWFUL EVIL

NEUTRAL EVIL

CHAOTIC EVIL
```
“...focusing on the wrong things"
- Erik Meijer
MODERN PROGRAMMING

- Experiments
- Cloud & Docker
- Code & Data
- Visualizations
Fifty Fizzbuzzes

An Extreme Programming Exercise by Vi Hart

One of the many things I learned from Extreme Programming is that working together is this technique for deeply exploring, releasing yourself from preconceptions, thinking in novel ways, and breaking down your most rigid structures. It's been very beneficial for me.

You make 50 of something; in one big mess, you put it all together, but it's important to exhaust your possibilities to break and illuminate habits, rather than merely making it a week.

50 is a LOT, and it takes you on this wild ride of learning, discovering and attempting and combining and taking apart. It can be oddly emotional and intense.

This technique was meant for artists, but now I've used it for Web stuff and programming language design too. So now I'm going to learn Python, over the weekend I made 50 implementations of Fizzbuzz.

Fizzbuzz is a classic simple programming exercise and common interview question based on a classic kids game, where you take turns counting, but if the number is divisible by 3 you say "fizz", and if it's divisible by 5 you say "buzz", and if it's divisible by both you say "Fizzbuzz".

- Ok, so this is a jupyter notebook, so if you want to run stuff you'll need to download this and open it, for which you might need https://www.anaconda.com/download/
- Luckily twitter user @paulib3r made a version that works online, go here and click Fifty Fizzbuzzes.ipynb:
  https://nbviewer.org/wn/https://vihart.com/channel/Fifty%20Fizzbuzzes.ipynb
- To run the examples, click in the box and press Ctrl+Enter:
- Each numbered section is independently runnable unless otherwise noted. For sections with multiple boxes you might have to run the parts above it for it to work.
- If things break or get weird, try kernel -> restart & clear output
- But things are going to get weird no matter what.
- Feel free to skip around. Some have visuals, some have sound. 48 is a complete text adventure.

1: The Obvious
Fifty Fizzbuzzes

An Extreme Programming Exercise by Vi Hart

One of the many things I learned from Evelyn Eastmond in my time working with her is this technique for deeply exploring, releasing yourself from preconceptions, thinking in new ways.

You make 50 of something. In one big marathon. (Maybe spaced over a few days, but it's important to exhaust your possibilities to break and illuminate habits, rather than making habits as with making a thing a week.)

50 is a LOT, and it takes you on this weird journey of ups and downs and complexifying and simplifying and combining and taking apart. It can be oddly emotional and intense.

This technique was meant for artists, but my research group has used it for VR stuff and programming language design too. So now to learn python, over the weekend I made 50 implementations of fizzbuzz.
JUPYTER IS OUR QUIPU

Representing narrative and data dynamically
PROBABILISTICALLY

programming

@lynnlangit
"OTHER PEOPLE DO THAT"

"DOESN'T WORK"

"TOO DIFFICULT"
Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

SAN FRANCISCO (Reuters) - Amazon.com Inc’s (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.
HOW TO BEGIN?
DIALECT

a particular form of a language that is peculiar to a specific region or social group

synonyms: regional language, local language, local speech, vernacular, idiom;
regionalisms, localisms;
informal lingo

• COMPUTING

a particular version of a programming language
ML is statistics, calculus, linear algebra and more math

for example...partial differential equations

\[ \frac{\partial u}{\partial t} = h^2 \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right) \]
def laplace(x):
    laplace_k = make_kernel([[0.5, 1.0, 0.5],
                             [1.0, -6., 1.0],
                             [0.5, 1.0, 0.5]])
    return simple_conv(x, laplace_k)

N = 500
u_init = np.zeros([N, N], dtype=np.float32)
ut_init = np.zeros([N, N], dtype=np.float32)

for n in range(40):
    a,b = np.random.randint(0, N, 2)
    u_init[a,b] = np.random.uniform()

eps = tf.placeholder(tf.float32, shape=())
damping = tf.placeholder(tf.float32, shape=())

U = tf.Variable(u_init)
Ut = tf.Variable(ut_init)
U_ = U + eps * Ut
Ut_ = Ut + eps * (laplace(U) - damping * Ut)

step = tf.group(
    U.assign(U_),
    Ut.assign(Ut_))

tf.global_variables_initializer().run()
VISUAL
DIALECT
ONE ML DIALECT?

Is this a CAT or DOG?
"EASY" TENSOR FLOW
LEVELING DIALECT

(input: every single pixel of the image)

- Are there patches of lines?
- Are there ovals?
- Are there triangles?
- Are there fur?
- Are there 2 eyes?
- Are there 2 ears?
- Is there a nose?

Is it a dog?

yes/no

© Machine Learning @ Berkeley

@lynnlangit
Tensorflow in 5 Minutes (tutorial)

692,930 views
PRACTICALLY programming
A TRANSLATION PROBLEM: LANGUAGE > MATH
USE
THE
CLASSICS
START WITH STATISTICS
The integral from 0 to infinity of e^{-st}a(t)dt = \mathcal{L}\{e^{at}\} = \frac{a}{s+a}
Training in progress...
Choose class-labeled data set. Say, "iris.tab" from documentation data sets.

Logistic Regression
Random Forest Classification
SVM

Cross-validation takes place here. Double click to see the performance scores.

Several learners can be scored in cross-validation at the same time.

It's always a good idea to check out the data first.

Select a cell in confusion matrix to obtain related data instances. Here we examine them in the spreadsheet.

Use for additional analysis of cross-validation results.

Learners
- Logistic Regression
- Random Forest Learner
- SVM Learner

Predicted
- Iris-setosa
- Iris-versicolor
- Iris-virginica

Actual
- Iris-setosa
- Iris-versicolor
- Iris-virginica

Show:
- Number of instances

Select
- Select Correct
- Select Misclassified
- Clear Selection

USE TOOLS
ORANGE
**FASTAI LIBRARIES AND...**

```python
code
path = untar_data(URLs.MNIST_SAMPLE)
data = ImageDataBunch.from_folder(path)
learn = ConvLearner(data, tvm.resnet18, metrics=accuracy)
learn.fit(1)
```

<table>
<thead>
<tr>
<th>epoch</th>
<th>train loss</th>
<th>valid loss</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.081393</td>
<td>0.046429</td>
<td>0.985770</td>
</tr>
</tbody>
</table>

- Training Videos
- Jupyter Notebooks
- Kaggle Kernels
- AWS Resources
# Kotlin for Data Science

<table>
<thead>
<tr>
<th>Math Expression</th>
<th>Kotlin Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sum_{i=1}^{3} i )</td>
<td>(1..3).sum()</td>
</tr>
<tr>
<td>( \sum_{i=1}^{5} 10i )</td>
<td>(1..5).map { i -&gt; 10*i }.sum()</td>
</tr>
<tr>
<td>( \sum_{i=0}^{n} i^2 )</td>
<td>fun f(n: Int) = (0..n).map { it.pow(2) }.sum()</td>
</tr>
<tr>
<td>( \sum_{i=1}^{100} 3x^2 + 2i )</td>
<td>fun f(x: Double) = (1..100).map { i -&gt; 3 * x.pow(2) + (2*i) }.sum()</td>
</tr>
<tr>
<td>( 10 + 3 \sum_{i=0}^{n} i^2 )</td>
<td>fun f(n: Int) = 10 + 3*(0..n).map { it * it }.sum()</td>
</tr>
<tr>
<td>( \sum_{i=0}^{n} x_i + i )</td>
<td>fun f(allX: List&lt;Int&gt;) = allX.mapIndexed { i, x -&gt; x + i }.sum()</td>
</tr>
<tr>
<td>( \sum_{i=1}^{4} \sum_{j=4}^{20} 2ij )</td>
<td>(1..4).flatMap { i -&gt; (4..20).map { j -&gt; 2 * i * j } }.sum()</td>
</tr>
<tr>
<td>( \prod_{i=1}^{n} i )</td>
<td>(1..n).fold(1L, Long::times)</td>
</tr>
</tbody>
</table>
COLLABORATION & LEARNING

From the course: Agile Software Development: Pair and Mob Programming

Watching: Scenario: "Hello World service" in Ballerina
KAGGLE IS THE NEW GITHUB
JUST USE XGBOOST

David Smith Retweeted

Szilard [Deeper than Deep Learning] @DataScienceLA · 2h

5/n So is deep learning going to wipe out all other methods in all these domains? Nope 🤷‍♂️Fonts 🤷‍♂️ 60% of Kaggle competitions have been won essentially with gradient boosting machines (xgboost) (+feature engineering +ensembles).

kaggle

Machine Learning Challenge Winning Solutions

- The most frequently used tool by data scientists is Python.
- The most popular libraries are pandas, numpy, and scikit-learn.
- A lot of contributions from the kaggle community.
GOOGLE BUYS KAGGLE
MSFT
BUYS
LOBE
C# Code Prediction with a Neural Network

Jul 20, 2018

**TLDR:** I used Python to create a neural network that implements an F# function to predict C# code. The network was compiled to a CoreML model and runs on iOS to be used in my app *Continuous* to provide keyboard suggestions.
Python Code Prediction

Next word/sequence prediction for Python code

INPUT

{"code": "for i in ", "num_results": 5}

OUTPUT

[
    [
        "range",
        0.6257643103599548
    ],
    [
        "<UNK>",
        0.1413002461194992
    ]
]
CODE IS DATA

Microsoft acquires GitHub for $7.5 billion
WHAT CONSTITUTES DATA?
# What Data Do You See?

## TAX INVOICE

**Original**

<table>
<thead>
<tr>
<th>LICENSEE</th>
<th>LICENSOR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name:</strong> Lynn Langit Consulting</td>
<td><strong>Name:</strong> Hub Australia</td>
</tr>
<tr>
<td><strong>Address:</strong></td>
<td><strong>Address:</strong> Level 2, 696 Bourke Street</td>
</tr>
<tr>
<td><strong>ABN:</strong></td>
<td><strong>Location:</strong> Melbourne, VIC, 3000</td>
</tr>
<tr>
<td><strong>Contact:</strong></td>
<td><strong>ABN:</strong> 45 145 858 304</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Product Name</th>
<th>Quantity</th>
<th>Unit Price</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hyde Park - Non-member Day Pass</td>
<td>1</td>
<td>50.00 AUD</td>
<td>50.00 AUD</td>
</tr>
</tbody>
</table>

**Payment Method**: Credit Card or Ezidebit via members.hubaustralia.com

**Subtotal**: 50.00 AUD
**Total**: 55.00 AUD
What Data Do You See?

TAX INVOICE

Original

LICENSSEE
Name: Lynn Langit Consulting
Address:
ABN:
Contact:

LICENSOR
Name: Hub Australia
Address: Level 2, 696 Bourke Street
Location: Melbourne, VIC, 3000
ABN: 45 145 858 304

# | Product Name              | Quantity | Unit Price | Amount  
---|--------------------------|----------|------------|---------
1  | Hyde Park - Non-member Day Pass | 1        | 50.00 AUD  | 50.00 AUD  

Payment Method: Credit Card or Ezidebit via members.hubaustralia.com
Subtotal: 50.00 AUD
GST: 5.00 AUD
Total: 55.00 AUD

@lynnlangit
# What Data Do I See?

## Tax Invoice

### Licensee
- **Name:** Lynn Langit Consulting
- **Address:**
- **ABN:**
- **Contact:**

### Licensor
- **Name:** Hub Australia
- **Address:** Level 2, 696 Bourke Street
- **Location:** Melbourne, VIC, 3000
- **ABN:** 45 145 858 304

### Invoice Details
- **No:** INV - 6988
- **Issue Date:** 06/09/2018
- **Due Date:** 06/09/2018

### Items
<table>
<thead>
<tr>
<th>#</th>
<th>Product Name</th>
<th>Quantity</th>
<th>Unit Price</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hyde Park - Non-member Day Pass</td>
<td>1</td>
<td>50.00 AUD</td>
<td>50.00 AUD</td>
</tr>
</tbody>
</table>

### Payment Method
- **Credit Card or EziDebit via members.hubaustralia.com**

### Total Charges
- **Subtotal:** 50.00 AUD
- **GST:** 5.00 AUD
- **Total:** 55.00 AUD
What Data Does He See?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>hub AUSTRALIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAX INVOICE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Licensee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name: Lynn Langit Consulting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABN:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Licensee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name: Hub Australia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address: Level 2, 696 Bourke Street</td>
<td></td>
<td></td>
</tr>
<tr>
<td>City: Melbourne, VIC, 3000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABN: 145 145 858 304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Product Name</td>
<td>Quantity</td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td>1</td>
<td>Hyde Park - Non-member Day Park</td>
<td>1</td>
</tr>
<tr>
<td>Payment Method</td>
<td>Credit card or Ezidebit via medipassaustralia.com</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>50.00 AUD</td>
<td></td>
</tr>
<tr>
<td>GST</td>
<td>5.00 AUD</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>55.00 AUD</td>
<td></td>
</tr>
</tbody>
</table>
BIAS
USING THEIR MODEL
GOOGLE "STATISTICS"
Introduction to ML Fairness

Faceting | X-Axis | Faceting | Y-Axis | Y-Axis | Display | Color | Display | Type | Position | X-Axis
---|---|---|---|---|---|---|---|---|---|---
age | age | gender | gender | education | native_country | <DEFAULT>
CAN YOU SEE BIAS?

---

What-If Tool demo - binary classifier for smile detection in images - CelebA dataset

Select a datapoint to get started

Clicking on a datapoint in the visualization will load all the features and values associated with that example. Some of the things you can do in the datapoint editor are:

- Test inference on edited values
- Edit features and run inference to see how your model performs

Compute distances from a selected datapoint
Have the selected example be an anchor and create a new distance feature for all loaded examples

Find closest counterfactuals
See the closest example with a different classification

Legend
Colors

Link

@lynnlangit
Data Science Ethics Checklist

A. Data Collection

- **A.1 Informed consent**: If there are human subjects, have they given informed consent, where subjects affirmatively opt-in and have a clear understanding of the data uses to which they consent?
- **A.2 Collection bias**: Have we considered sources of bias that could be introduced during data collection and survey design and taken steps to mitigate those?
- **A.3 Limit PII exposure**: Have we considered ways to minimize exposure of personally identifiable information (PII) for example through anonymization or not collecting information that isn’t relevant for analysis?

B. Data Storage

- **B.1 Data security**: Do we have a plan to protect and secure data (e.g., encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)?
- **B.2 Right to be forgotten**: Do we have a mechanism through which an individual can request their personal information be removed?
- **B.3 Data retention plan**: Is there a schedule or plan to delete the data after it is no longer needed?

C. Analysis

- **C.1 Missing perspectives**: Have we sought to address blindspots in the analysis through engagement with relevant stakeholders (e.g., checking assumptions and discussing implications with affected communities and subject matter experts)?
- **C.2 Dataset bias**: Have we examined the data for possible sources of bias and taken steps to mitigate or address these biases (e.g., stereotype perpetuation, confirmation bias, imbalanced classes, or omitted confounding variables)?
- **C.3 Honest representation**: Are our visualizations, summary statistics, and reports designed to honestly represent the underlying data?
- **C.4 Privacy in analysis**: Have we ensured that data with PII are not used or displayed unless necessary for the analysis?
- **C.5 Auditability**: Is the process of generating the analysis well documented and audit if necessary as it is the future?

D. Modeling

- **D.1 Proxy discrimination**: Have we ensured that the model does not rely on variables or proxies for variables that are unfairly discriminatory?
- **D.2 Fairness across groups**: Have we tested model results for fairness with respect to different affected groups (e.g., tested for disparate error rates)?
- **D.3 Metric selection**: Have we considered the effects of optimizing for our defined metrics and considered additional metrics?
- **D.4 Explainability**: Can we explain in understandable terms a decision the model made in cases where a justification is needed?
- **D.5 Communicate bias**: Have we communicated the shortcomings, limitations, and biases of the model to relevant stakeholders in ways that can be generally understood?

E. Deployment

- **E.1 Redress**: Have we discussed with our organization a plan for response if users are harmed by the results (e.g., how does the data science team evaluate these cases and update analysis and models to prevent future harm)?
- **E.2 Roll back**: Is there a way to turn off or roll back the model in production if necessary?
- **E.3 Concept drift**: Do we test and monitor for concept drift to ensure the model remains fair over time?
- **E.4 Unintended use**: Have we taken steps to identify and prevent unintended uses and abuse of the model and do we have a plan to monitor these once the model is deployed?
“THEY ARE THE SAME”

“IT IS NUANCED”

“THEY ARE DIFFERENT”

BIAS...ERROR
POSSIBILITES

programming
The most important data - our DNA
BY 2050 ~ 50% SEQUENCED
BIOINFORMATICS
CODING TO SUPPORT CANCER RESEARCH
VariantSpark / src / main / scala / au / csiro / variantspark / algo /

- metrics
- CanSplitTypes.scala: code cleaning and fix typos in variable names (#3)
- DecisionTree.scala: Fix: Make results reproducible (#86)
- PairwiseOperation.scala: Add API for Pairwise Operations (#50), 11 months ago
- PredictiveModel.scala: code cleaning and fix typos in variable names (#3), a year ago
- RandomForest.scala: Make -rre (randomize equal) option enabled by default (#36), a year ago
- WideKMeans.scala: adding ScalaDocs to ML files, a year ago
- package.scala: Switch importance to work on bytes, 2 years ago

Latest commit fb95fae on Jun 26
The plot above shows that VariantSpark has recovered the correct genotypes of this multivariate phenotype with interacting features (multiplicative and additive effects).

1. **chr2_223034082** (rs2218065) encoding for monobrow is the most important feature
2. A group of SNPs encoding for the MEGF10 gene (**chr5_126626044**), which is involved in Retina horizontal cell formation as the second most important marker, explaining why hipsters prefer checked shirts
3. **chr7_17284577** (rs4410790) the marker for increased coffee consumption is ranked third
4. **chr4_54511913** (rs4864809) the marker for beards is fourth

The last two are in swapped order compared to the formular of the HipsterIndex, however with 0.5 and 1 as weight they may be difficult to differentiate.
“HOW SHOULD WE SCALE IT?”
JUST

DOCKER

IT
```ballerina
import ballerina/http;
import ballerinax/docker;
import ballerinax/kubernetes;

@docker:Config {}

@docker:Expose()

@kubernetes:Ingress {}

@kubernetes:Service {}

@kubernetes:Deployment {}

endpoint http:Listener airlineEP { port:9091; }

@final string ECONOMY = "Economy";
@final string BUSINESS = "Business";
@final string FIRST = "First";

@http:ServiceConfig {basePath: "/airline"}
service<http:Service> airlineReservationService bind airlineEP {

    @http:ResourceConfig {methods: ["POST"], path: "/reserve", consumes: ["application/json"],...} reserveTicket(endpoint client, http:Request request) {...

}
```
Ballerina

Ballerina is a cloud-native programming language that incorporates fundamental concepts of distributed system integration into the language and offers a type-safe, concurrent environment to implement microservices with distributed transactions, reliable messaging, stream processing, and workflows.

About Ballerina

Ballerina is a compiled, transactional, statically and strongly typed programming language with textual and graphical syntaxes. Ballerina incorporates fundamental concepts of distributed system integration into the language and offers a type-safe, concurrent environment to implement microservices with distributed transactions, reliable messaging, stream processing, and workflows.

Ballerina is a language designed to be integration simple. Based around the interactions of sequence diagrams, Ballerina has built-in support for common integration patterns and connectors, including distributed transactions, compensation and circuit breakers. With first-class support for JSON and XML, Ballerina makes it simple and effective to build robust integration across network endpoints.
BUILD
ANALYSIS
PIPELINES
PIPELINE = NOTEBOOK, ML++

Genomic Sequence Files
- vcf
- fastq
- ...

Collect

Prepare
- λ-driven
- λ-portal
- λ-compress
- λ-convert

Process

Machine Learning - GWAS

Elastic Cluster

Data Lake

View/Use
- Web Tools
- Viewers
- Local Tools
- Jupyter Notebook(s)

@lynnlangit
CAN YOU SEE IT?
Scaling Custom Machine Learning on AWS

Understanding the Challenge

Bioinformatics is one of the most interesting and challenging areas to work on scaling big data machine learning solutions. These challenges include not only the size and scale of genomic data (3 Billion DNA ‘letters’ per person). They also include the potential to improve feedback loops for important research in human health, such as understanding significant variants in genomic data for potential CRISPR-Cas9 research. This research can have profound impact on diseases such as cancer.
CODE

for a better world
• True Cost of Execution Environment
• Data quality / features
• ML Model Hyperparameters
We become what we behold. We shape our tools, and then our tools shape us.

Marshall McLuhan
//TODO

- Check your Biases
- Learn more Statistics, Calculus & Linear Algebra
- Code with leveled Machine Learning
- Code with Subject Matter Experts
- Cloud configuration code is key to usability
- Visualize everything
TO PROGRAM

**WRITE** a series of coded software instructions **CONTROL** the operation of a computer or another machine
TO PROGRAM

CREATE a series of coded software instructions

GUIDE the operation of a computing environment
program
probably