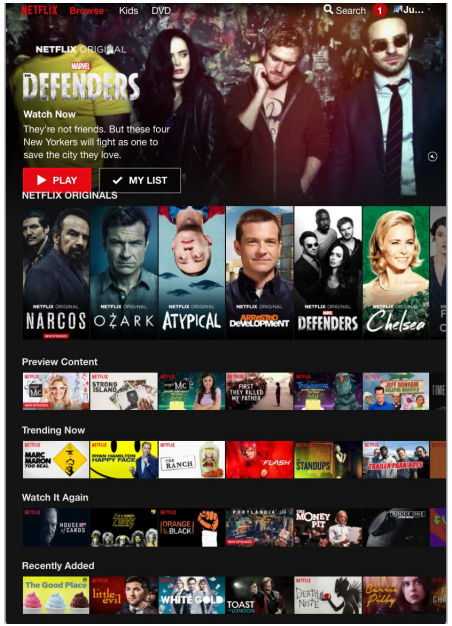
Vegas

The Missing Matplotlib for Scala/Spark

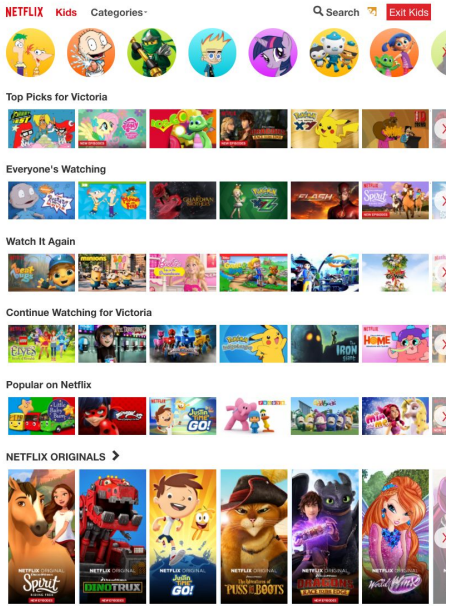
DB Tsai
Roger Menezes

NETFLIX

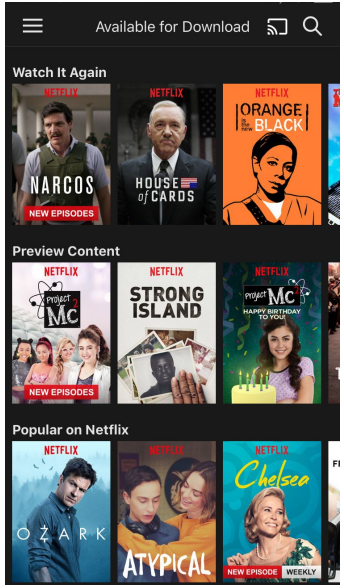
Netflix Recommendations



Homepage



Kids Page



Downloads Page

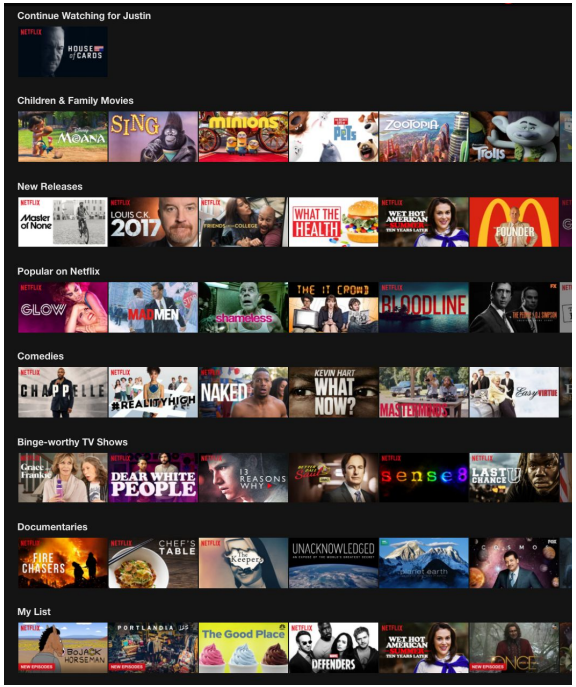
Every aspect of the Experience is Machine Learned

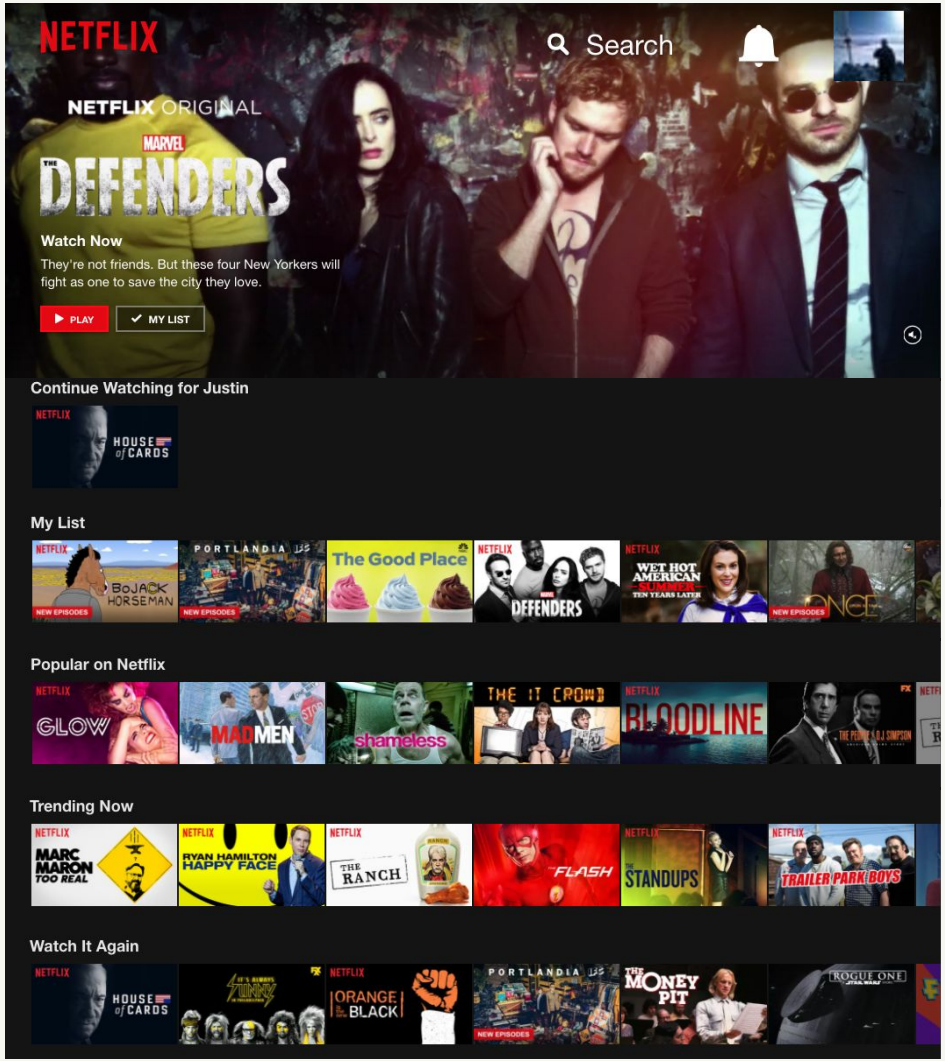
2017

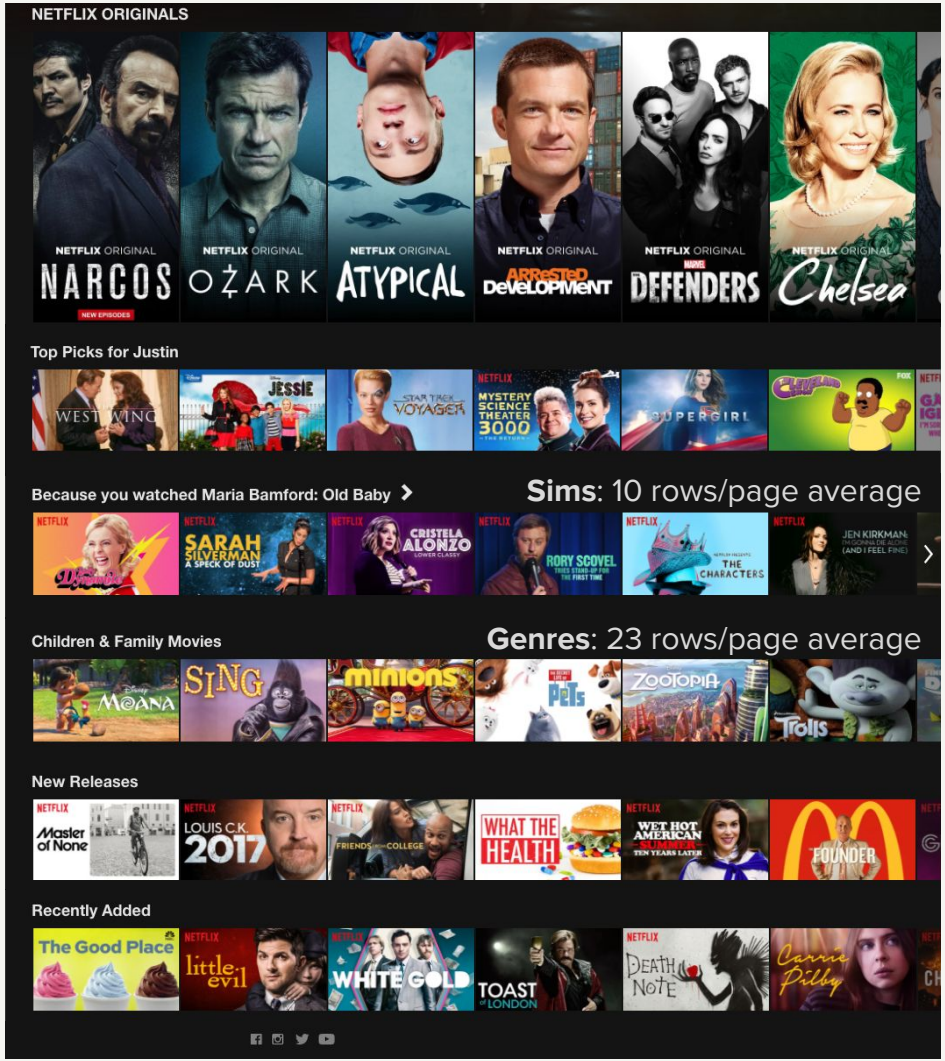
> 100M members

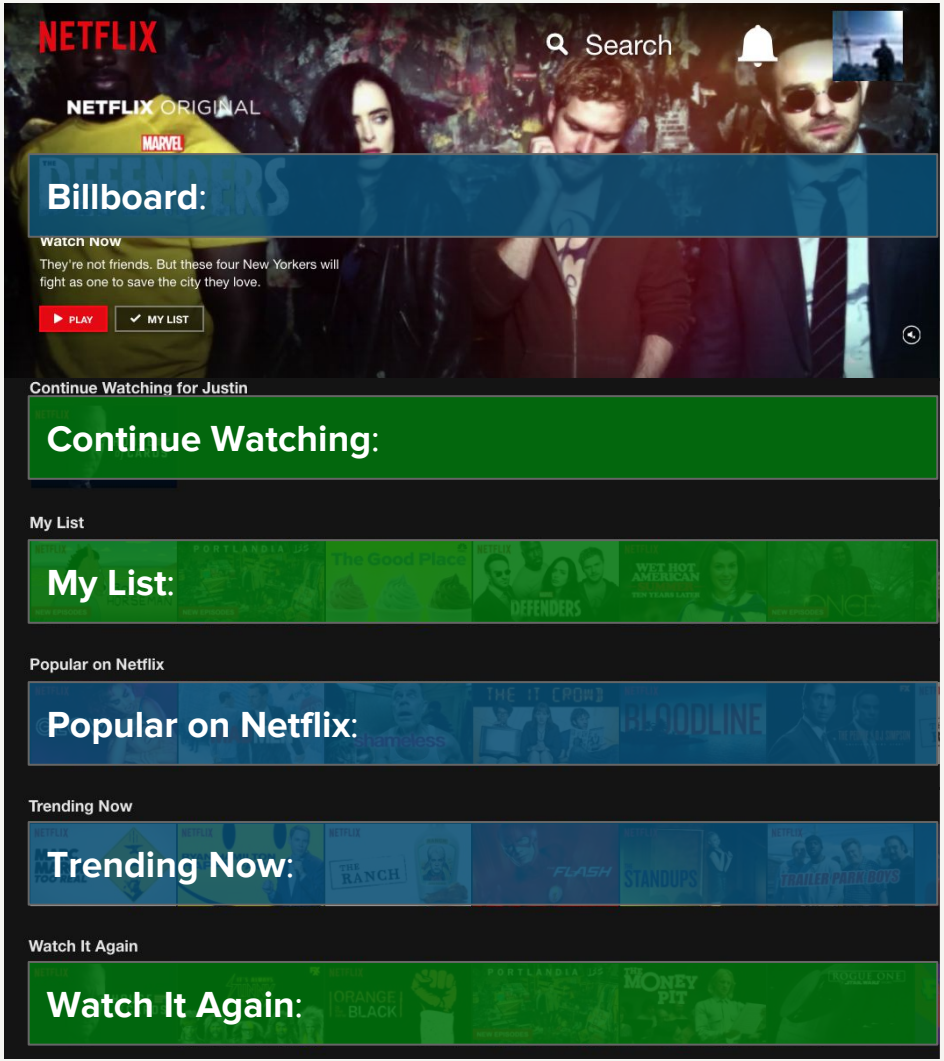
> 190 countries

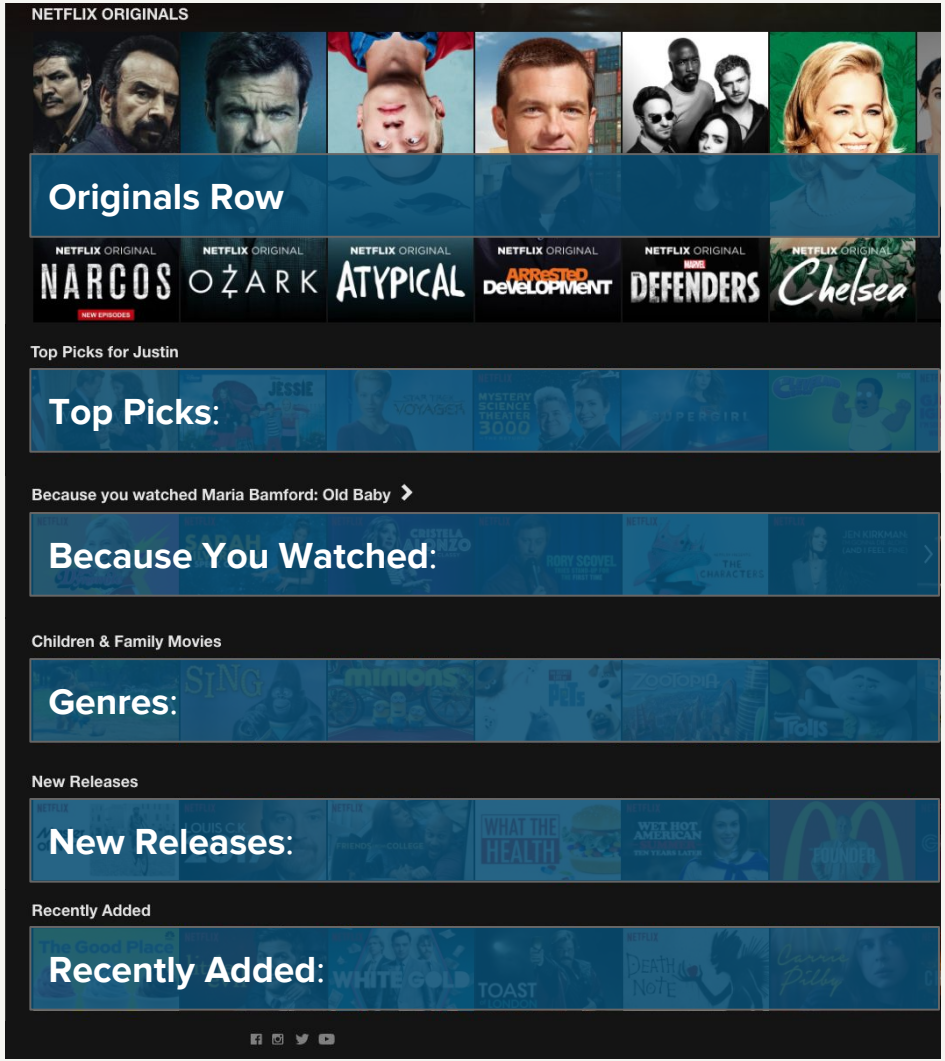
Multiple Devices




NETFLIX
Search
NETFLIX ORIGINAL
MARVEL
THE DEFENDERS
Watch Now
They're not friends. But these four New Yorkers will fight as one to save the city they love.
PLAY
MY LIST
Continue Watching for Justin
HOUSE of CARDS
My List
BoJACK HORSEMAN
PORTLANDIA
The Good Place
DEFENDERS
WET HOT AMERICAN SUMMER TEN YEARS LATER
ONCE
Popular on Netflix
GLOW
MAD MEN
shameless
THE IT CROWD
BLOODLINE
Trending Now
MARC MARON TOO REAL
RYAN HAMILTON HAPPY FACE
THE RANCH
FLASH
STANDUPS
TRAILER PARK BOYS
Watch It Again
HOUSE of CARDS
ORANGE is the BLACK
PORTLANDIA
THE MONEY PIT
ROGUE ONE


NETFLIX ORIGINALS
NETFLIX ORIGINAL NARCOS
NETFLIX ORIGINAL OZARK
NETFLIX ORIGINAL ATYPICAL
NETFLIX ORIGINAL ARRESTED DEVELOPMENT
NETFLIX ORIGINAL DEFENDERS
NETFLIX ORIGINAL Chelsea
Top Picks for Justin
WEST WING
JESSIE
STAR TREK VOYAGER
MYSTERY SCIENCE THEATER 3000
SUPERGIRL
Because you watched Maria Bamford: Old Baby
Sims: 10 rows/page average
Dreams
SARAH SILVERMAN A SPECK OF DUST
CRISTELA ALONZO
RORY SCOVEL
THE CHARACTERS
JEN KIRKMAN
Children & Family Movies
Genres: 23 rows/page average
MOANA
SING
MINIONS
PETS
ZOOTOPIA
Trolls
New Releases
Master of None
LOUIS C.K. 2017
FRIENDS from COLLEGE
WHAT THE HEALTH
WET HOT AMERICAN SUMMER TEN YEARS LATER
FOUNDER
Recently Added
The Good Place
little evil
WHITE GOLD
TOAST of LONDON
DEATH NOTE
Carrie Pilby

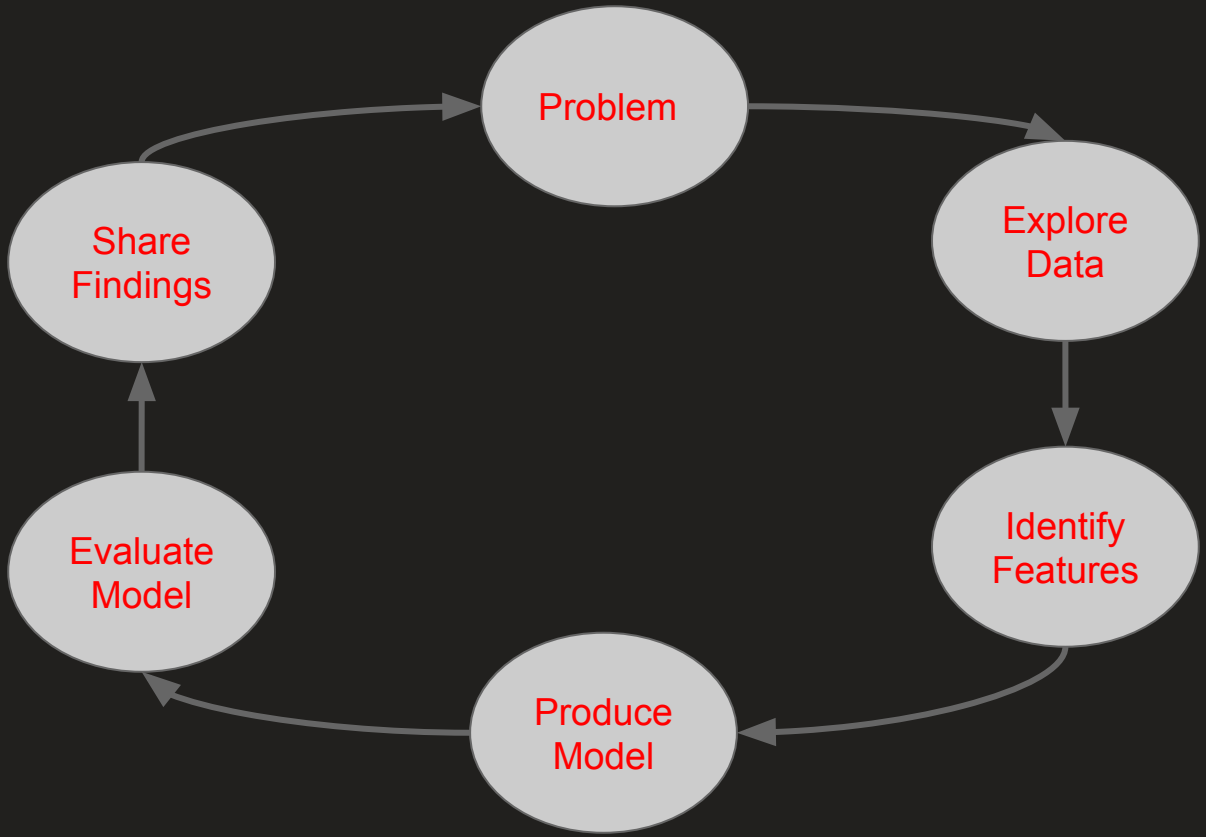

NETFLIX
Search
NETFLIX ORIGINAL
MARVEL
Billboard:
Watch Now
They're not friends. But these four New Yorkers will fight as one to save the city they love.
PLAY
MY LIST
Continue Watching for Justin
Continue Watching:
My List
My List:
Popular on Netflix
Popular on Netflix:
Trending Now
Trending Now:
Watch It Again
Watch It Again:


NETFLIX ORIGINALS
Originals Row
NETFLIX ORIGINAL NARCOS
NETFLIX ORIGINAL OZARK
NETFLIX ORIGINAL ATYPICAL
NETFLIX ORIGINAL ARRESTED DEVELOPMENT
NETFLIX ORIGINAL DEFENDERS
NETFLIX ORIGINAL Chelsea
Top Picks for Justin
Top Picks:
Because you watched Maria Bamford: Old Baby
Because You Watched:
Children & Family Movies
Genres:
New Releases
New Releases:
Recently Added
Recently Added:

Machine Learning at Netflix

- Optimize the Experimentation usecase vs Productionization
- Experimentation
  - Opportunity sizing, Data Exploration
  - Feature Identification and Selection
  - Tweaks to ML algos
  - Model Evaluation

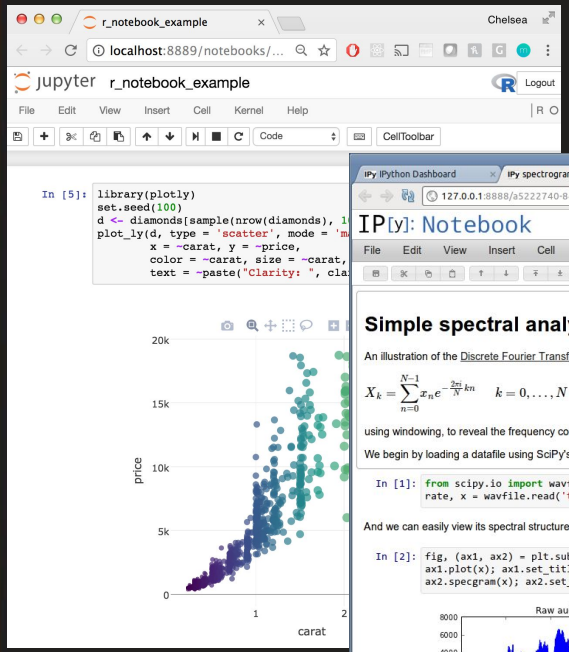
Experimenter's loop

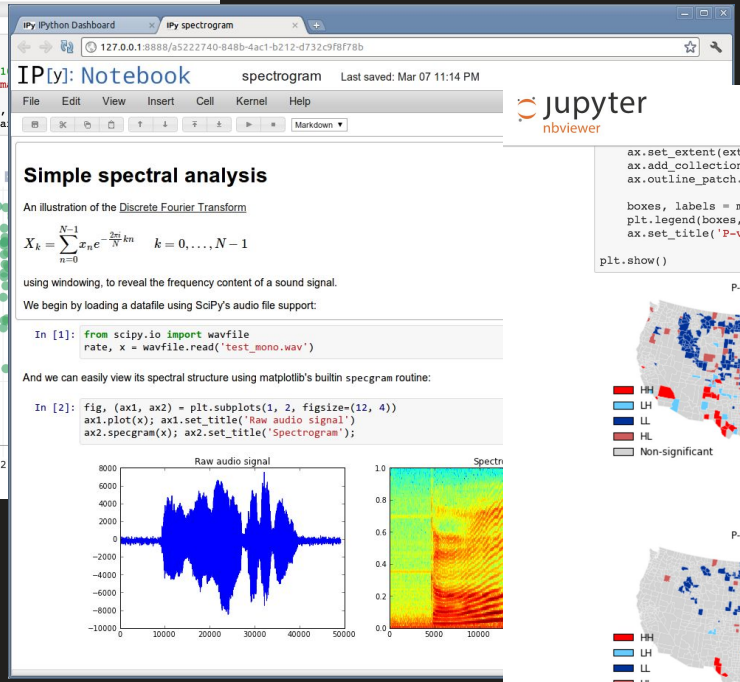


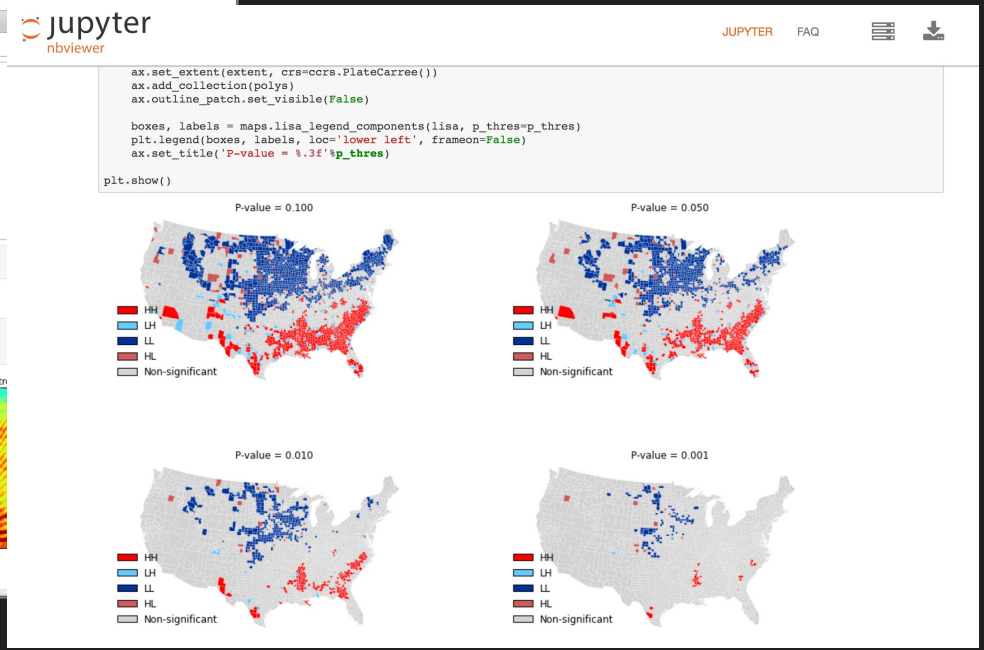
Notebooks

- Optimal for Experimentation
- Sharing reproducible research
  - Facilitates feedback loop with Product Managers
- End to end ML experiment.
  - Interactivity drives productivity

Python Notebooks







Python Notebooks

- Seamless Experience - ML experimentation
- Well known Scientific computing libraries
- Huge catalog of Visualization plotting libraries
    - Matplotlib, Seaborn, Bokeh, BQPlot, Lightning, etc.

Scala Notebooks

- Zeppelin, Jupyter, Databricks, Spark-Notebooks, ...
- Computing library gap filling up
- Lack of Visualization Libraries
  - Main friction point in adoption
  - End to End ML use case not convincing

Introducing Vegas

- Visualization Library in Scala
- Mainly built for the notebook use case
- Scala wrapper around Vega-Lite
    - Missing MatPlotLib for the Scala/Spark world.

NETFLIX

VEGAS

DECLARATIVE

STATISTICAL

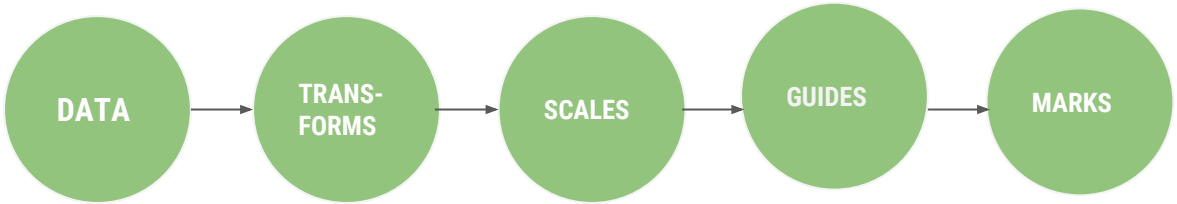
VISUALIZATION GRAMMAR

IN SCALA

You tell it ***WHAT*** should be done with the data, and it knows ***HOW*** to do it!

Operations such as ***filtering***, ***aggregation***, ***faceting*** are built into the visualization, rather than putting the burden on the user to massage the data into shape.

Complex visualizations can be built with a few high level abstractions:

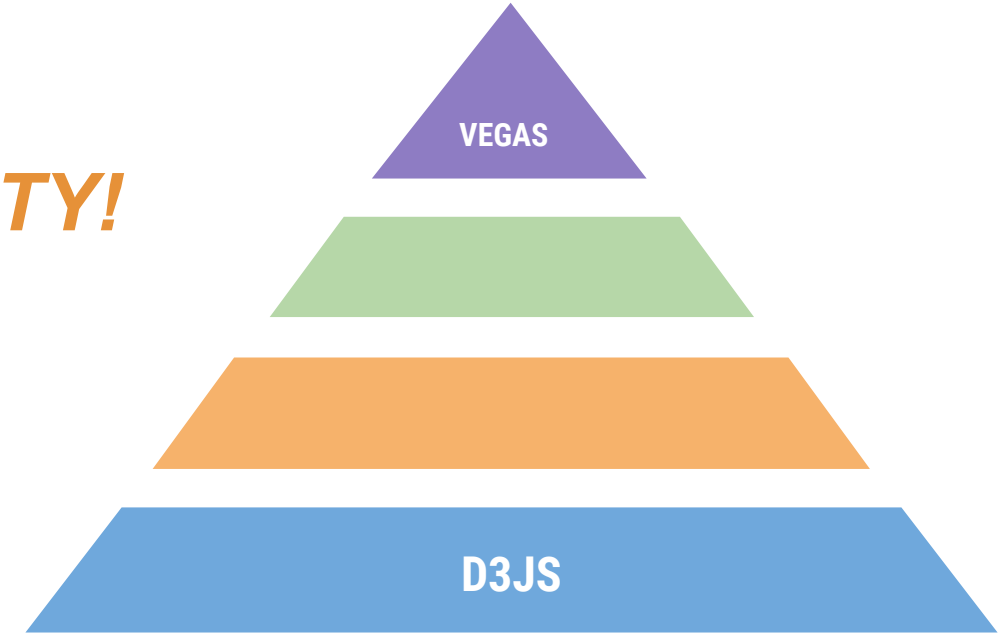


cf : Altair Talk by Brian Granger in PyData 2016 https://youtu.be/v5mrwq7yJc4

NETFLIX

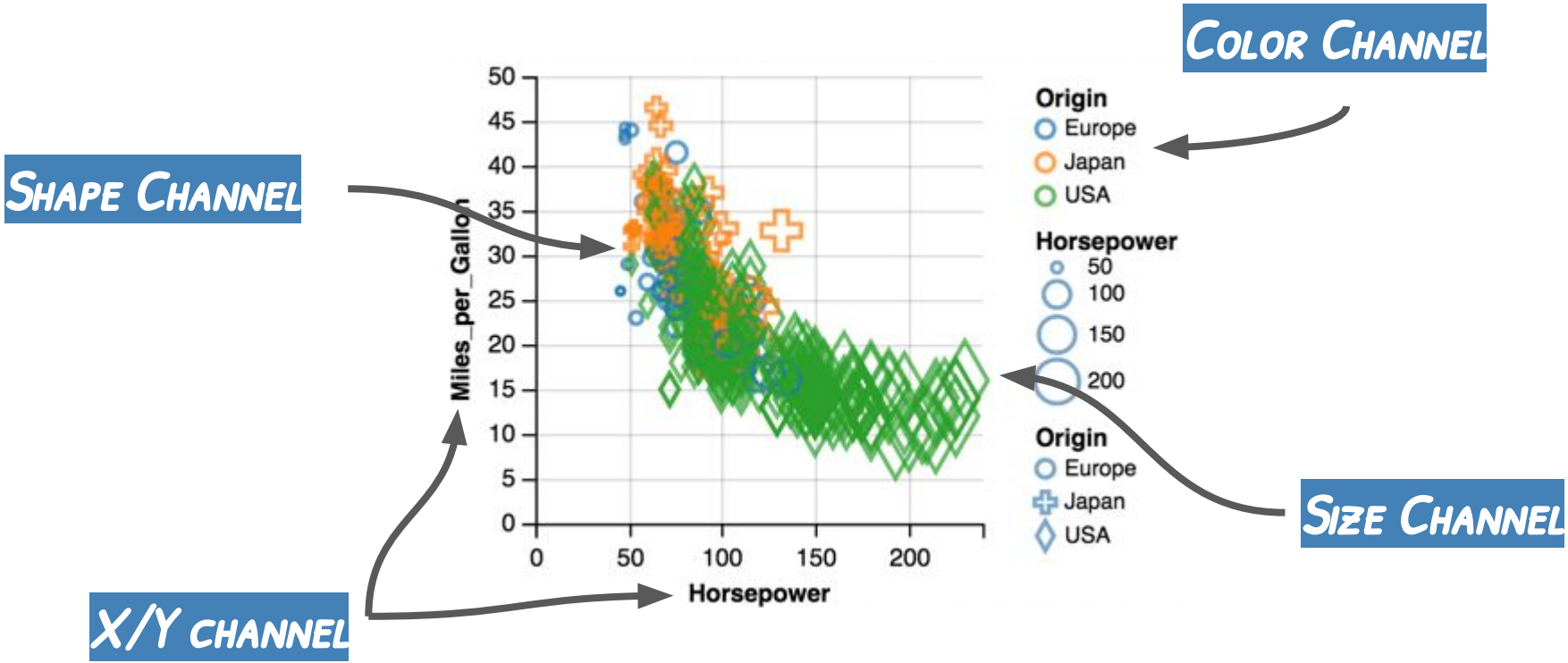
Added Bonus of Declarative Visualizations:

INTERACTIVITY!



VEGAS CODE EXPANDS OUT TO D3JS CODE!

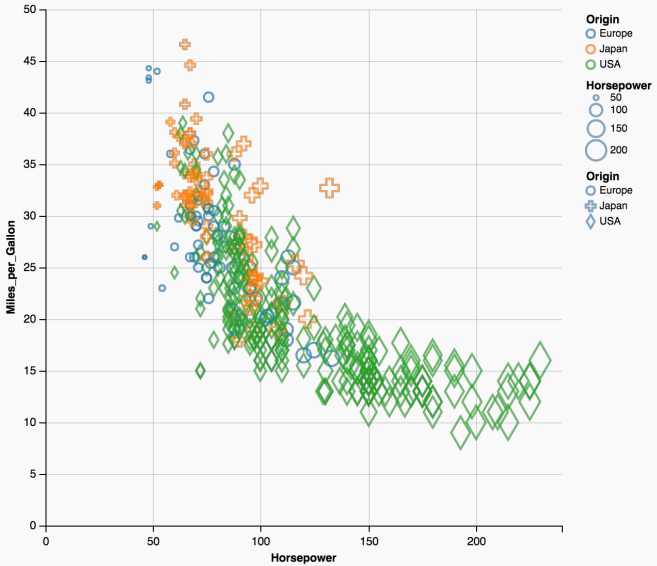
Anatomy of a plot: Channels



```
1  carsDf.printSchema
```

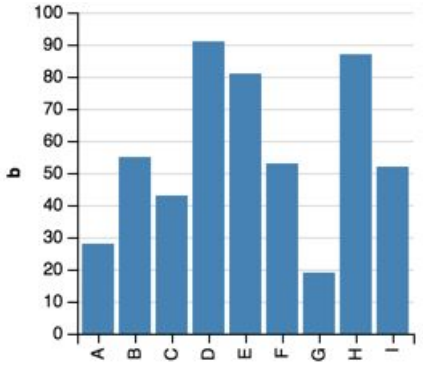
```
root
 |-- Acceleration: double (nullable = true)
 |-- Cylinders: long (nullable = true)
 |-- Displacement: double (nullable = true)
 |-- Horsepower: long (nullable = true)
 |-- Miles_per_Gallon: double (nullable = true)
 |-- Name: string (nullable = true)
 |-- Origin: string (nullable = true)
 |-- Weight_in_lbs: long (nullable = true)
 |-- Year: string (nullable = true)
```

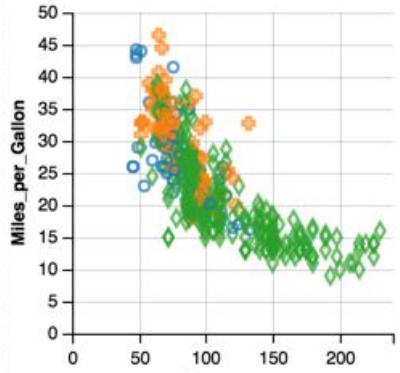
```
Vegas("Miles_per_Gallon vs horsepower").
  withDataFrame(carsDf).
  mark(Point).
  encodeX("Horsepower", Quantitative).
  encodeY("Miles_per_Gallon", Quantitative).
  encodeColor("Origin", Nominal).
  encodeSize("Horsepower", Quantitative).
  encodeShape("Origin", Nominal).
show
```

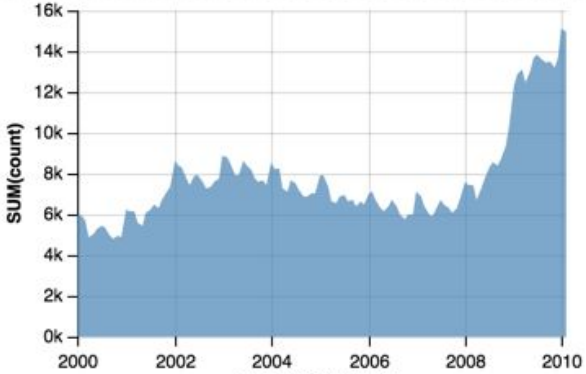


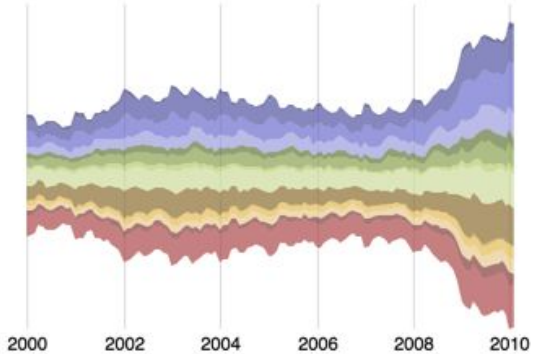
Features...

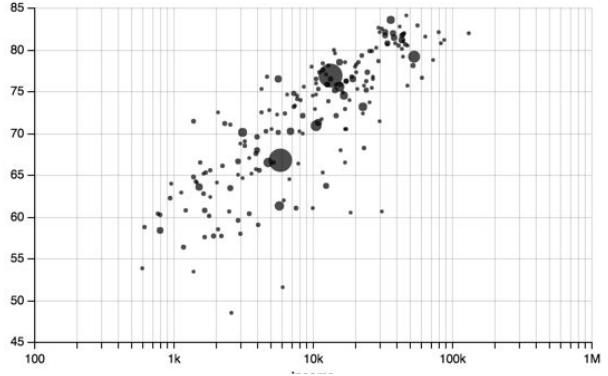
1. Supports most plot types



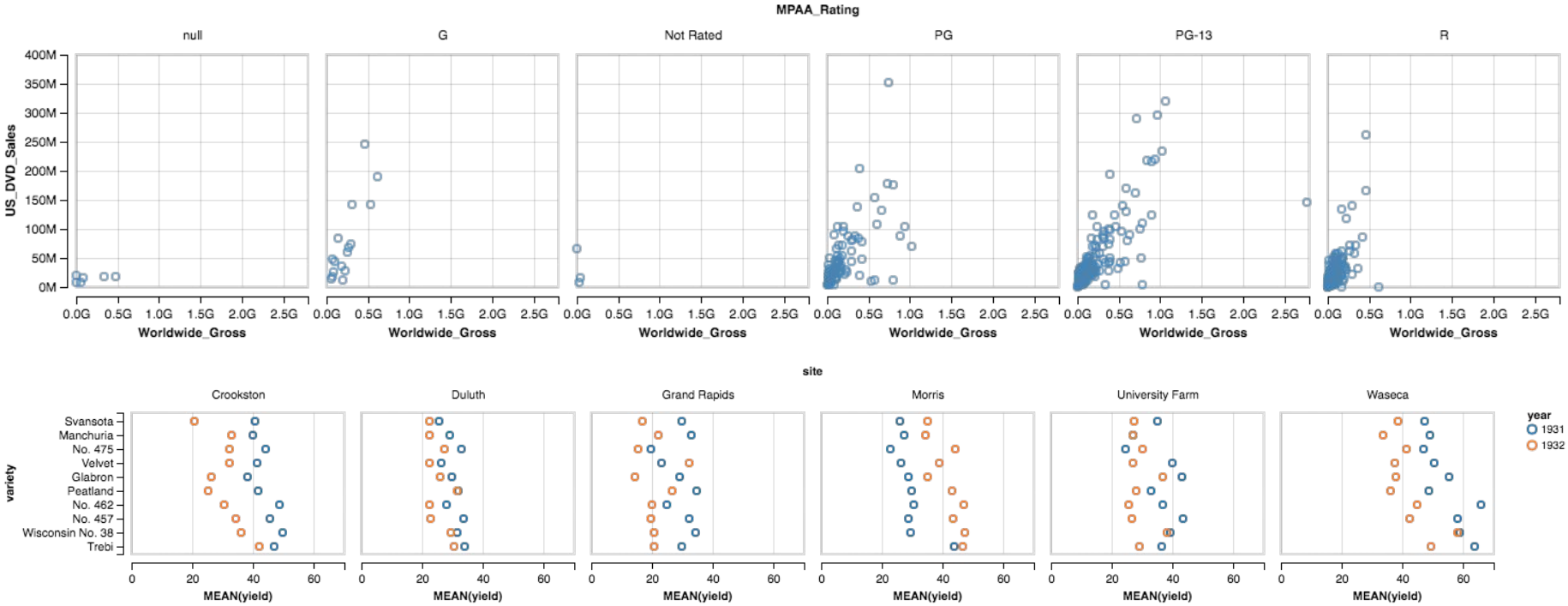




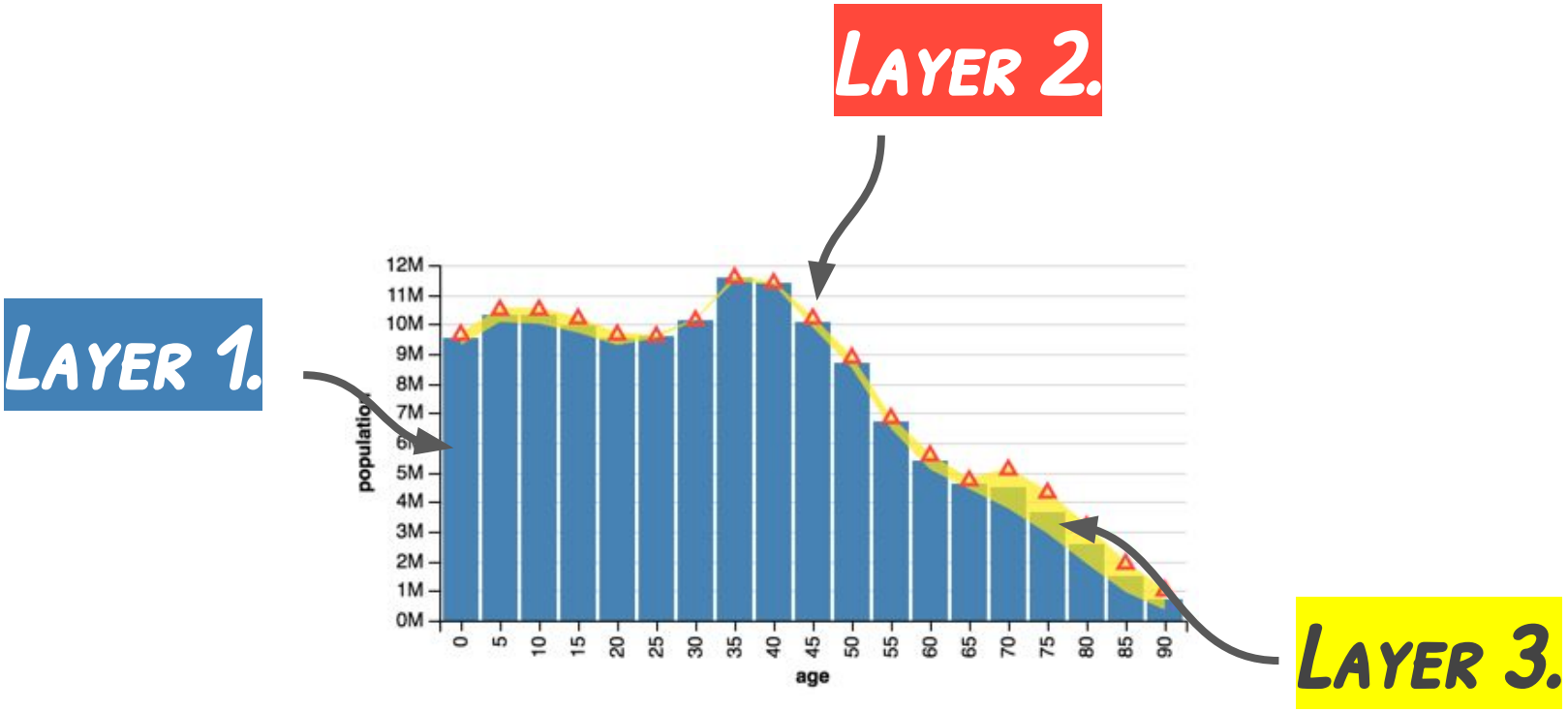




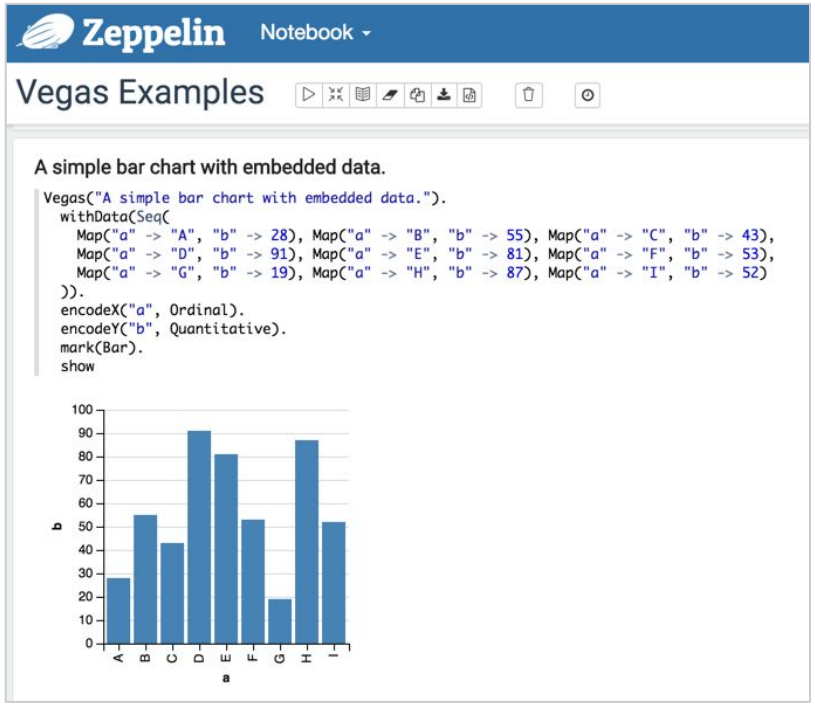
2. *Trellis plots*

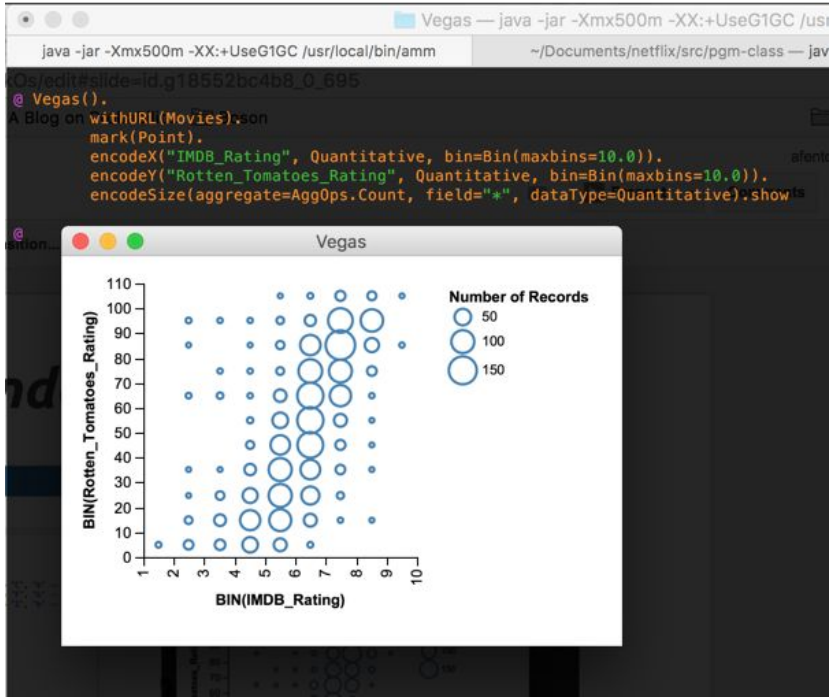


3. Layers

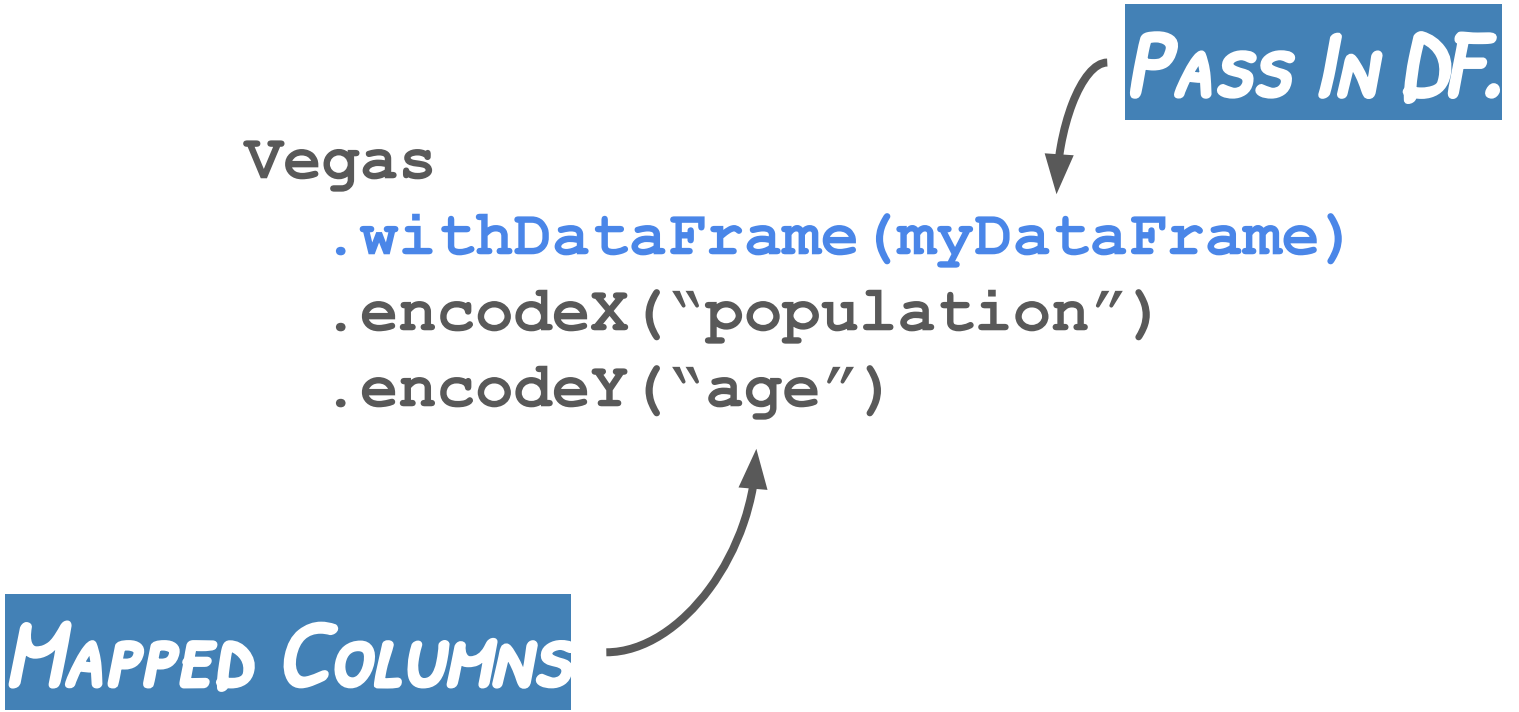

Layer 2.
Layer 1.
12M
11M
10M
9M
8M
7M
6M
5M
4M
3M
2M
1M
0M
population
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
age
Layer 3.

4. Notebook and Consoles





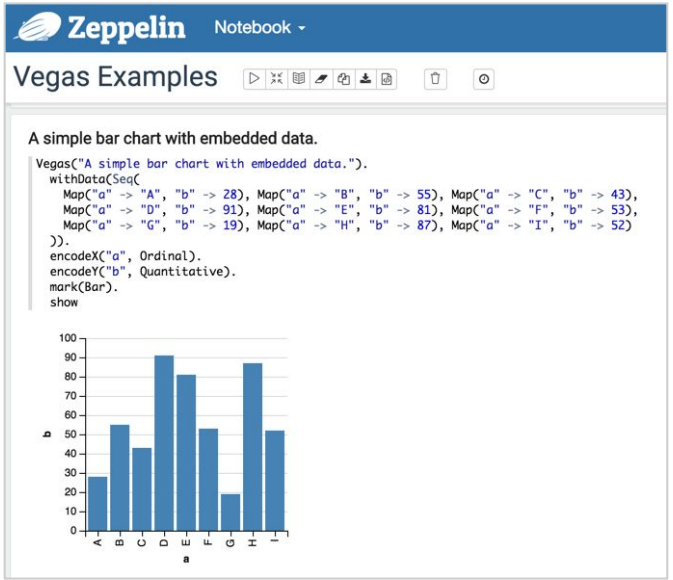
5. Built-in spark support



6. Visual statistics

- Advanced Binning
- Sorting
- Scaling
- Custom Transforms
- Time Series
- Aggregation
- Filtering
- Math functions (log, etc)
- Descriptive Statistics

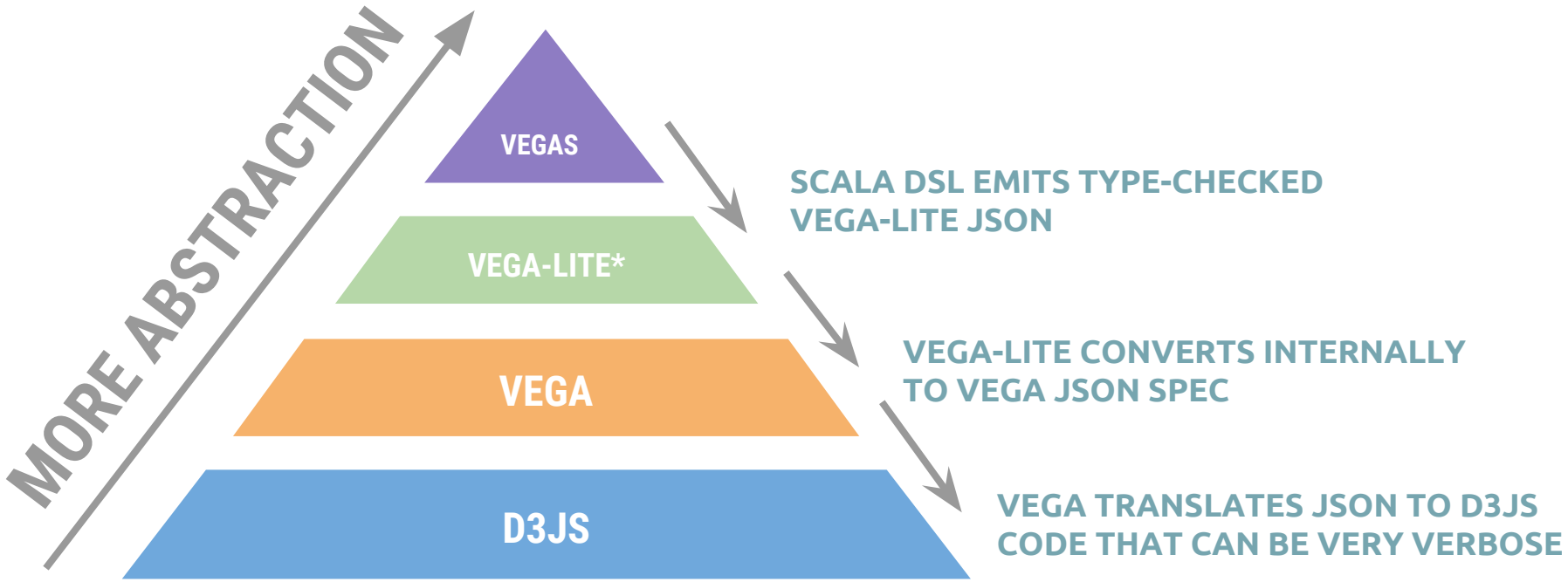
How It Works !



1. Specify in Scala
2. Embed HTML (iFrame)
3. Render within iFrame using JS

NETFLIX

A SCALA DSL FOR VEGA-LITE



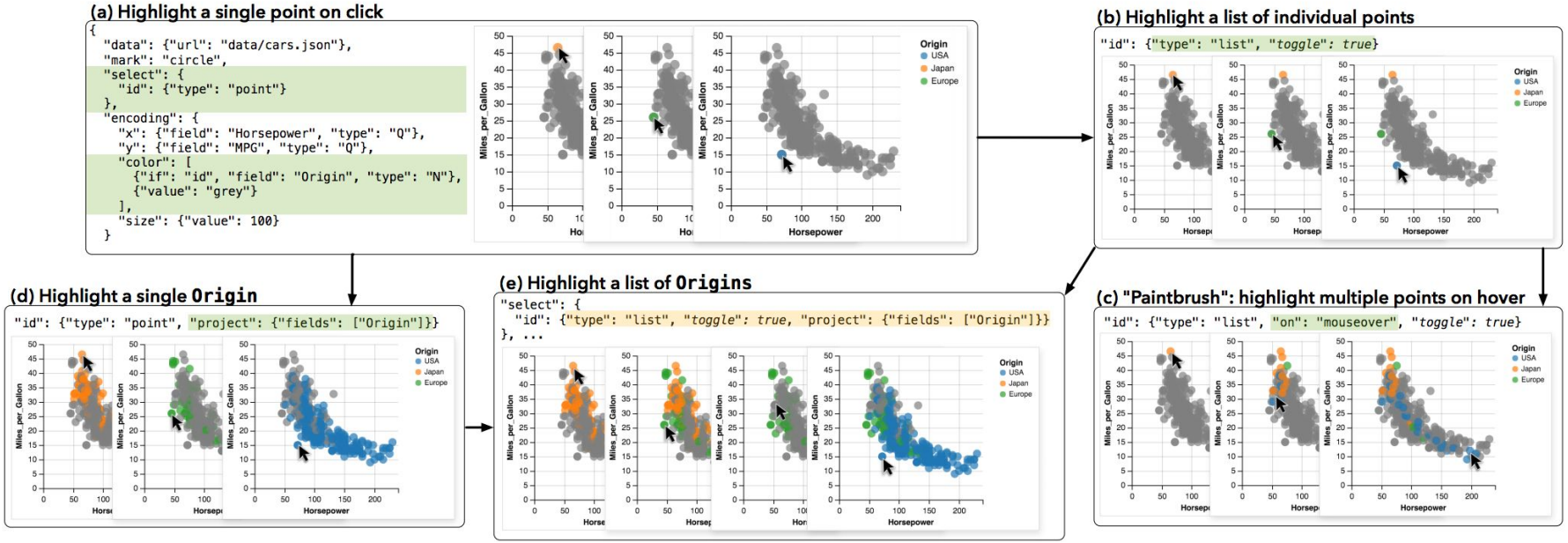
* *Vega-Lite*

Example 3

Other Channels + Transforms

What's coming

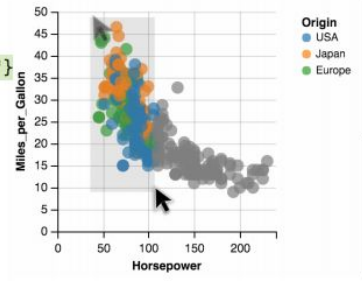
1. Interactive selections



2. Selections transforms

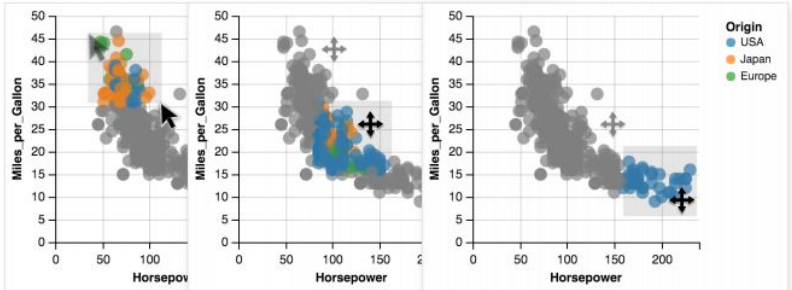
(a) Rectangular brush

```
"select": {
  "region": {"type": "interval"}
},
...
  "color": [
    {"if": "region", ...}
  ]
...
```



(b) Moving the brush

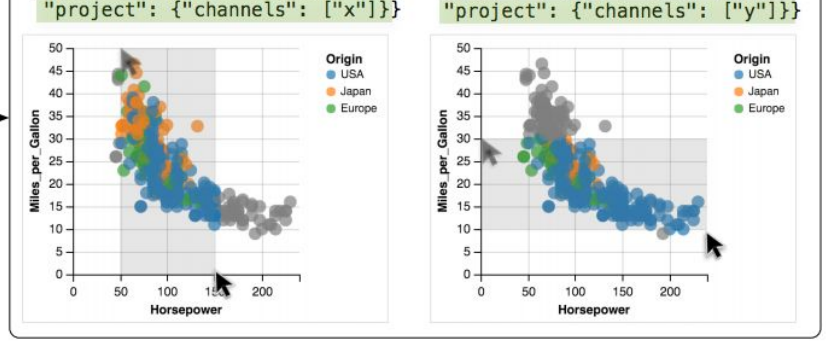
```
"region": {"type": "interval", "translate": true}
```



(c) Single-dimension brush

```
"region": {"type": "interval",
 "project": {"channels": ["x"]}}
```

```
"region": {"type": "interval",
 "project": {"channels": ["y"]}}
```



Contributors

Aish

DB

Roger

NETFLIX

Sudeep

Jeremy

Thank you.

NETFLIX

The missing MatPlotLib for Scala/Spark

http://vegas-viz.org

@NetflixResearch
@rogermenezes @dbtsai

NETFLIX