AI+VR: The future of data analytics

Ciro Donalek
Co-Founder & CTO

Aakash Indurkhya
Head of ML Projects

VIRTUALITICS™
immerse yourself in your data

NVIDIA GTC Conference
March, 18th 2019

Virtualitics fast facts

- Software company merging AI and VR to provide a novel way to find insights in data
- Founded in August 2016, based on 10+ years of research at Caltech and NASA/JPL in machine learning and immersive visualization
- Significant customer traction with DOD and Fortune 500 companies

Immersive analytics in VIP – Virtualitics Immersive Platform™

TechConnect DEFENSE INNOVATION AWARD 2018

1ST PLACE WINNER OpenWERX INTERNATIONAL DATA SCIENCE COMPETITION 2017

FinTech Innovation Lab

Technology trends - 2019

Trend No. 2: Augmented analytics

Data scientists now have increasing amounts of data to prepare, analyze and group — and from which to draw conclusions. Given the amount of data, exploring all possibilities becomes impossible. This means businesses can miss key insights from hypotheses the data scientists don't have the capacit to explore.

>> too much data, even for data scientist!

Trend No. 3: AI-driven development

AI-driven development looks at tools, technologies and best practices for embedding AI into applications and using AI to create AI-powered tools for the development process.

Trend No. 6: Immersive technologies

Through 2028, conversational platforms, which change how users interact with the world, and technologies such as augmented reality (AR), mixed reality (MR) and virtual reality (VR), which change how users perceive the world, will lead to a new immersive experience. AR, MR and VR show potential for increased productivity, with the next generation of VR able to sense shapes and track a user's position and MR enabling people to view and interact with their world.

Technology trends - 2019

Trend No. 2: Augmented analytics

Data scientists now have increasing amounts of data to prepare, analyze and group — and from which to draw conclusions. Given the amount of data, exploring all possibilities becomes impossible. This means businesses can miss key insights from hypotheses the data scientists don't have the capacit to explore.

>> too much data, even for data scientist!

Trend No. 3: AI-driven development

AI-driven development looks at tools, technologies and best practices for embedding AI into applications and using AI to create AI-powered tools for the development process.

>> AI, more AI, but also API!

Trend No. 6: Immersive technologies

Through 2028, conversational platforms, which change how users interact with the world, and technologies such as augmented reality (AR), mixed reality (MR) and virtual reality (VR), which change how users perceive the world, will lead to a new immersive experience. AR, MR and VR show potential for increased productivity, with the next generation of VR able to sense shapes and track a user's position and MR enabling people to view and interact with their world.

Technology trends - 2019

Trend No. 2: Augmented analytics

Data scientists now have increasing amounts of data to prepare, analyze and group — and from which to draw conclusions. Given the amount of data, exploring all possibilities becomes impossible. This means businesses can miss key insights from hypotheses the data scientists don't have the capacit to explore.

>> too much data, even for data scientist!

Trend No. 3: AI-driven development

AI-driven development looks at tools, technologies and best practices for embedding AI into applications and using AI to create AI-powered tools for the development process.

>> AI, more AI, but also API!

Trend No. 6: Immersive technologies

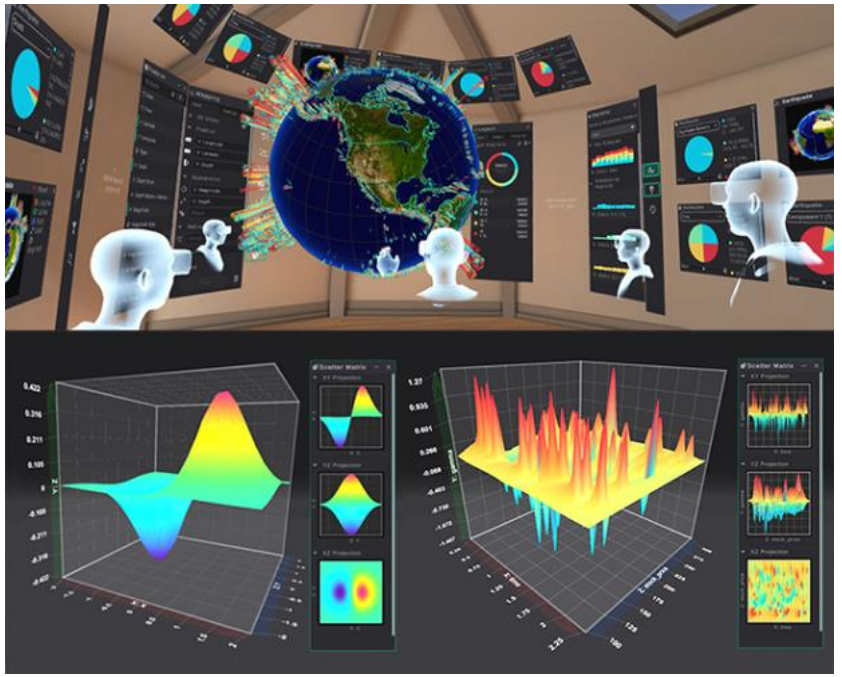
Through 2028, conversational platforms, which change how users interact with the world, and technologies such as augmented reality (AR), mixed reality (MR) and virtual reality (VR), which change how users perceive the world, will lead to a new immersive experience. AR, MR and VR show potential for increased productivity, with the next generation of VR able to sense shapes and track user's position and MR enabling people to view and interact

>> VR is here to stay!

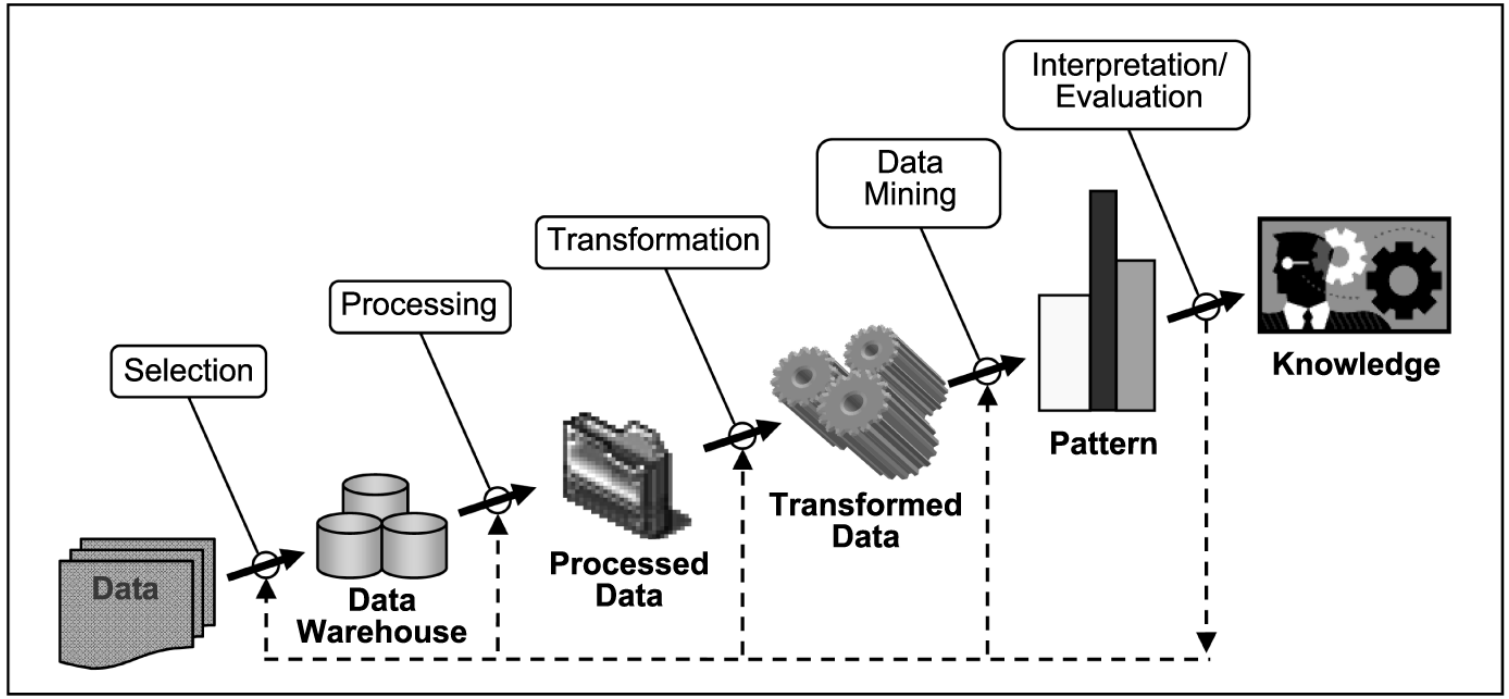
The importance of data visualization

“If I can’t picture it, I can’t understand it”

- Key in understanding data
- Visual access to huge amount of data
- Helps understand not only the data but also how it was collected
- Humans have an incredible pattern recognition system, but don’t scale!

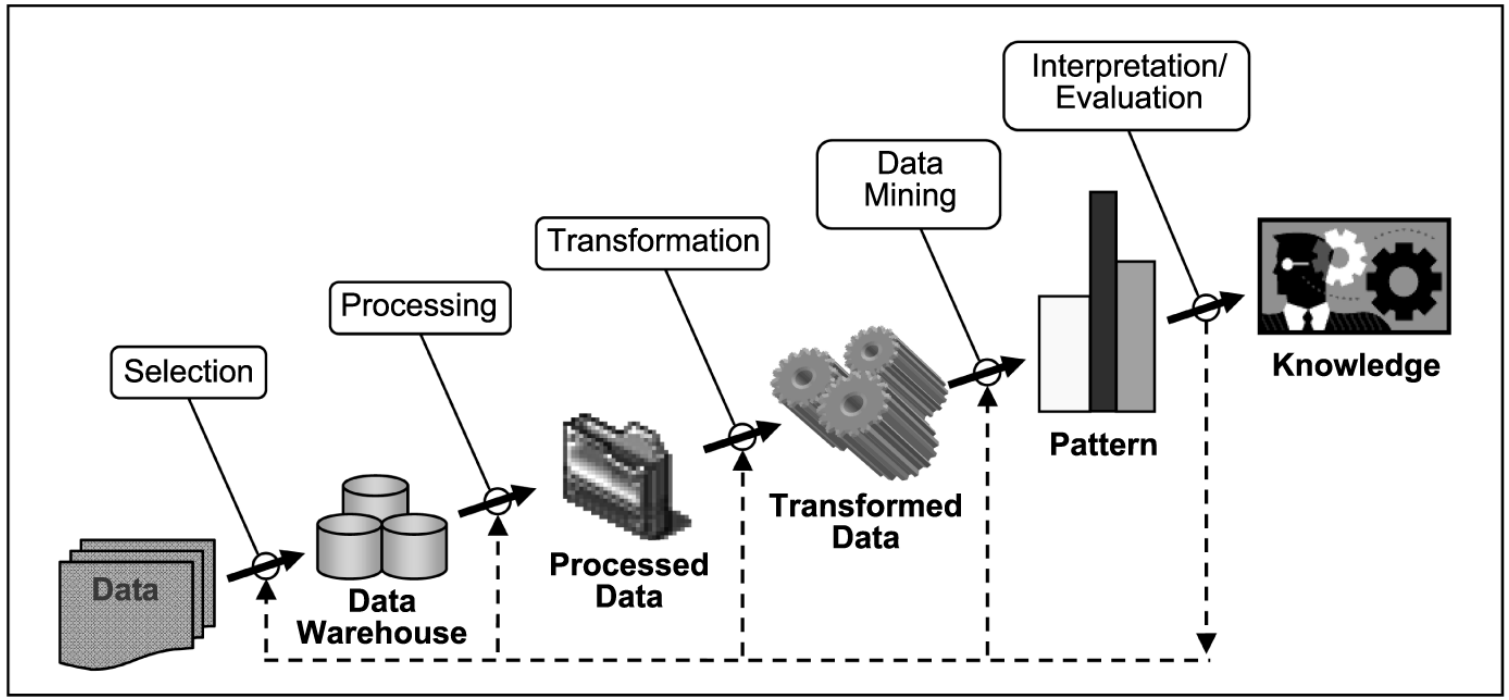


KDD: visualization crucial at every step

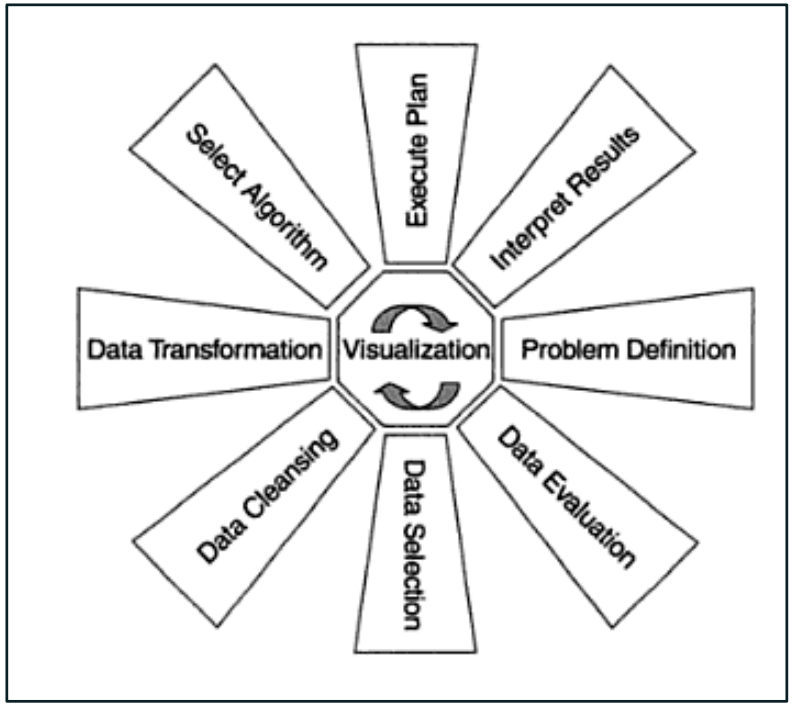


Knowledge Discovery in Databases

KDD: visualization crucial at every step



Knowledge Discovery in Databases



Visualization-centric view:
visualization and interaction essential at every step.

Data has evolved, BI tools haven't

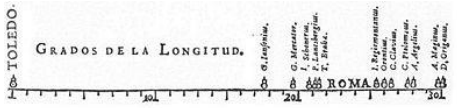
Evolution of data

- Complex: Hundreds of metrics for each sample
- From static to dynamic, from centralized to distribute resources, from data sets to data streams
- From "next day" analysis to real-time

Not easy to find insights with standard tools

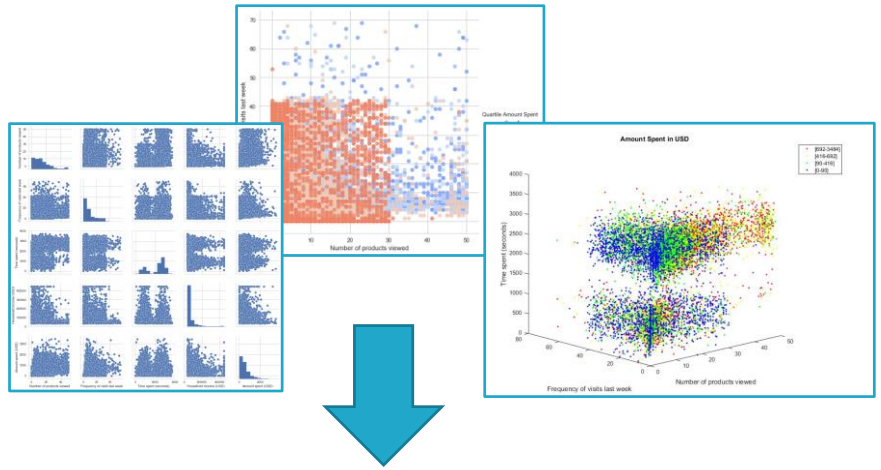
"Big Data is not a revolution, but an evolution"

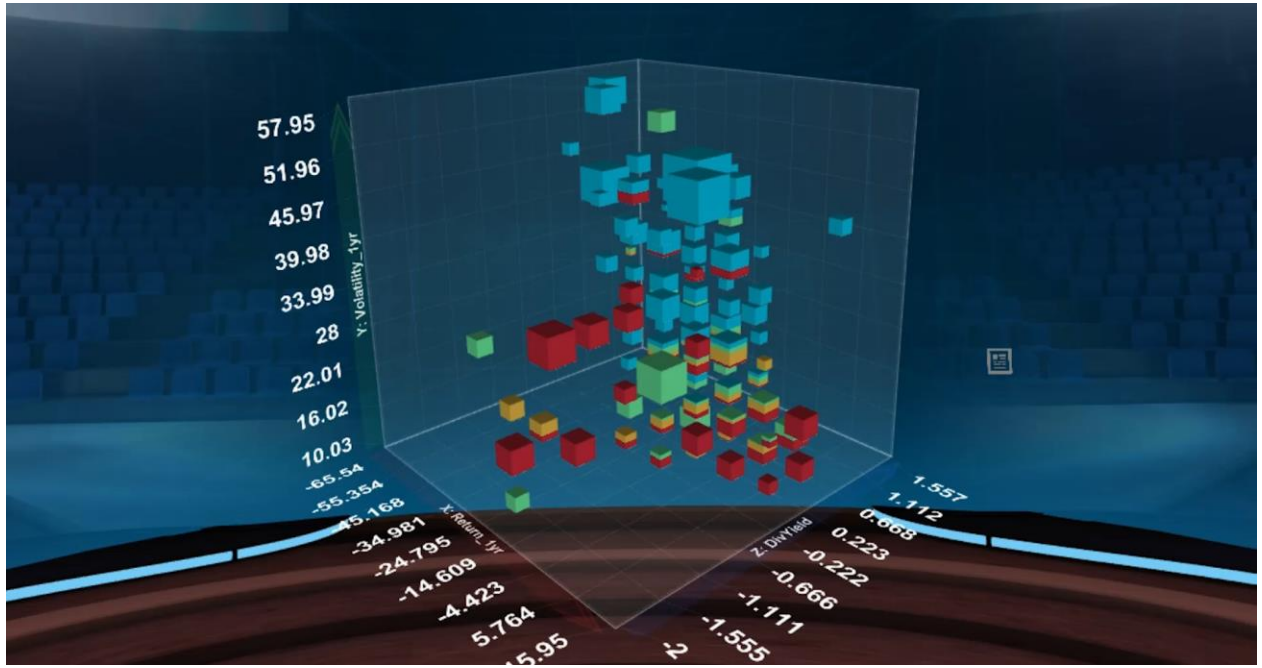
Evolution of data visualization

From Van Langren 

to VR

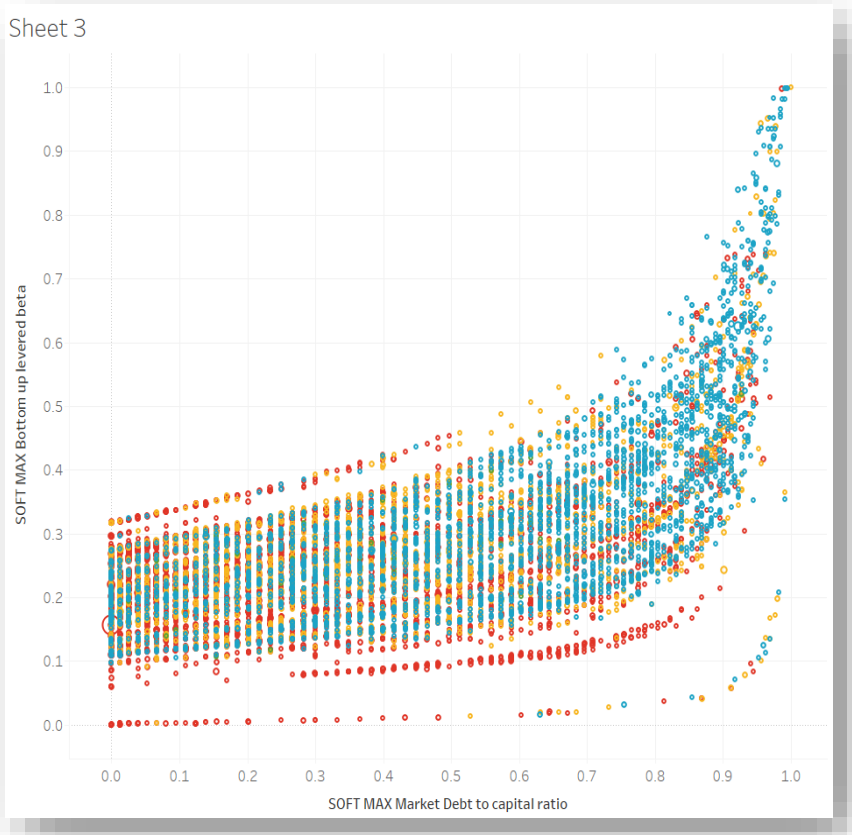
- Standard visualization tools not enough anymore
- We need to re-imagine the way we represent and interact with data to get key insights
- Clever interaction makes it easier to look for patterns
- Virtual Assistants and AI-driven visualization

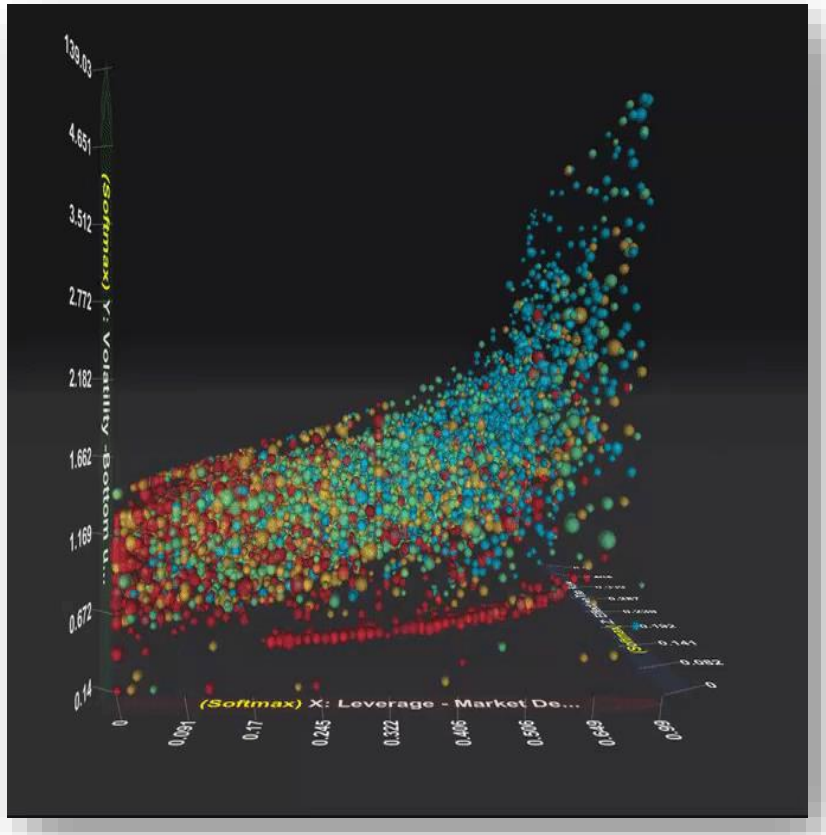




From 2D to 3D and… beyond!

- Traditional 2D data visualization may hide crucial information
- Beyond 3D: new ways to encode, visualize and understand more dimensions



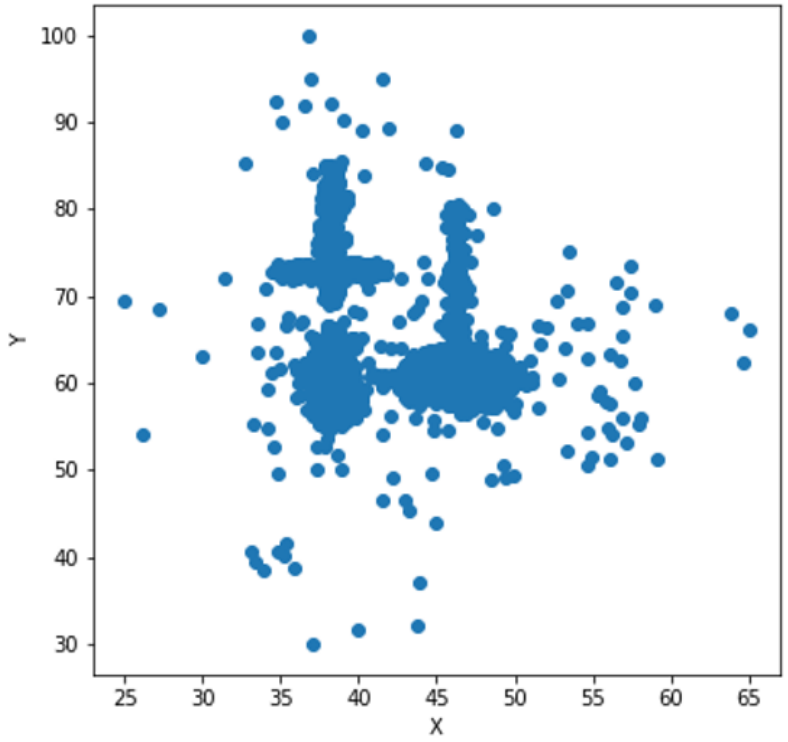


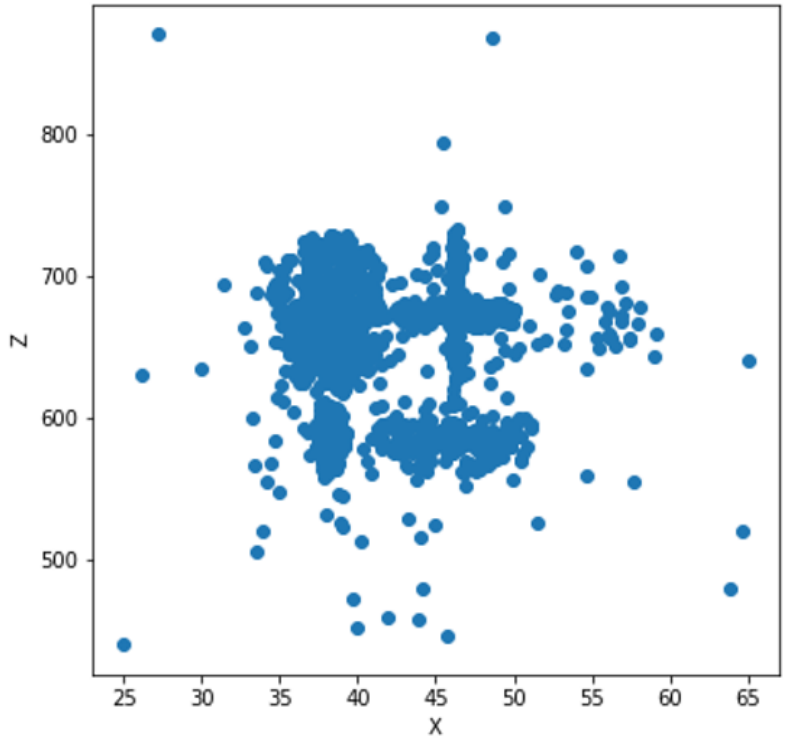
Dimensions Encoded in VIP

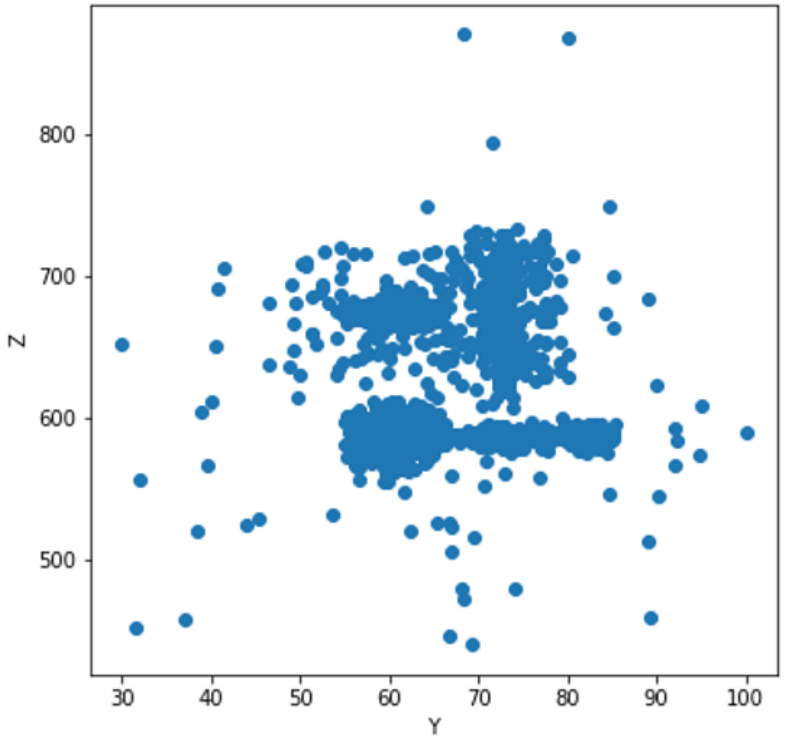
Position (X, Y, Z)
Colors, Opacity
Shape, Texture
Size, Halo, Vectors
Pulsation, Sounds
Spinning, Glow, Sounds

New class of “3D” problems

Standard 2D plots may hide crucial information visible only at the intersection of many dimensions

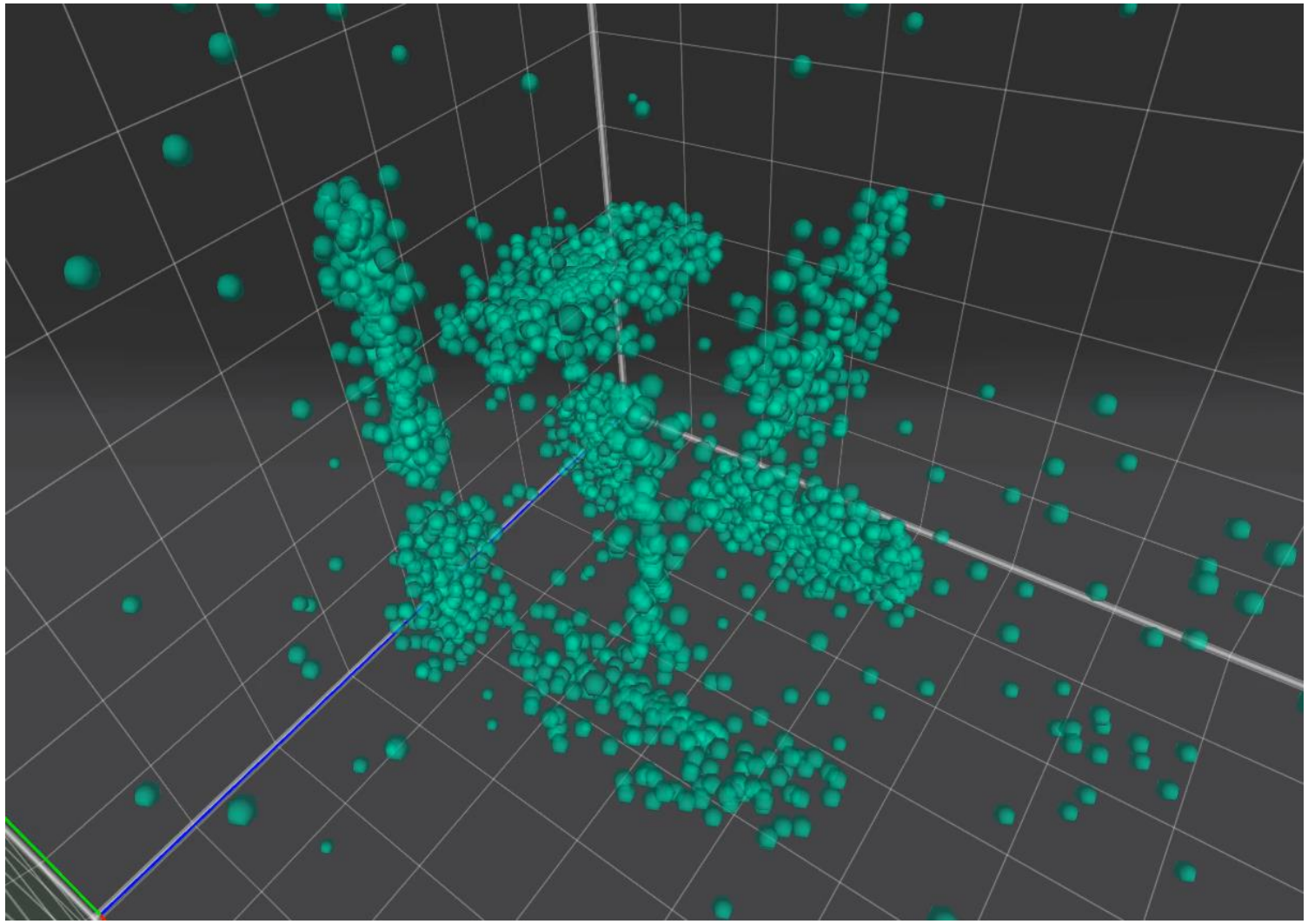






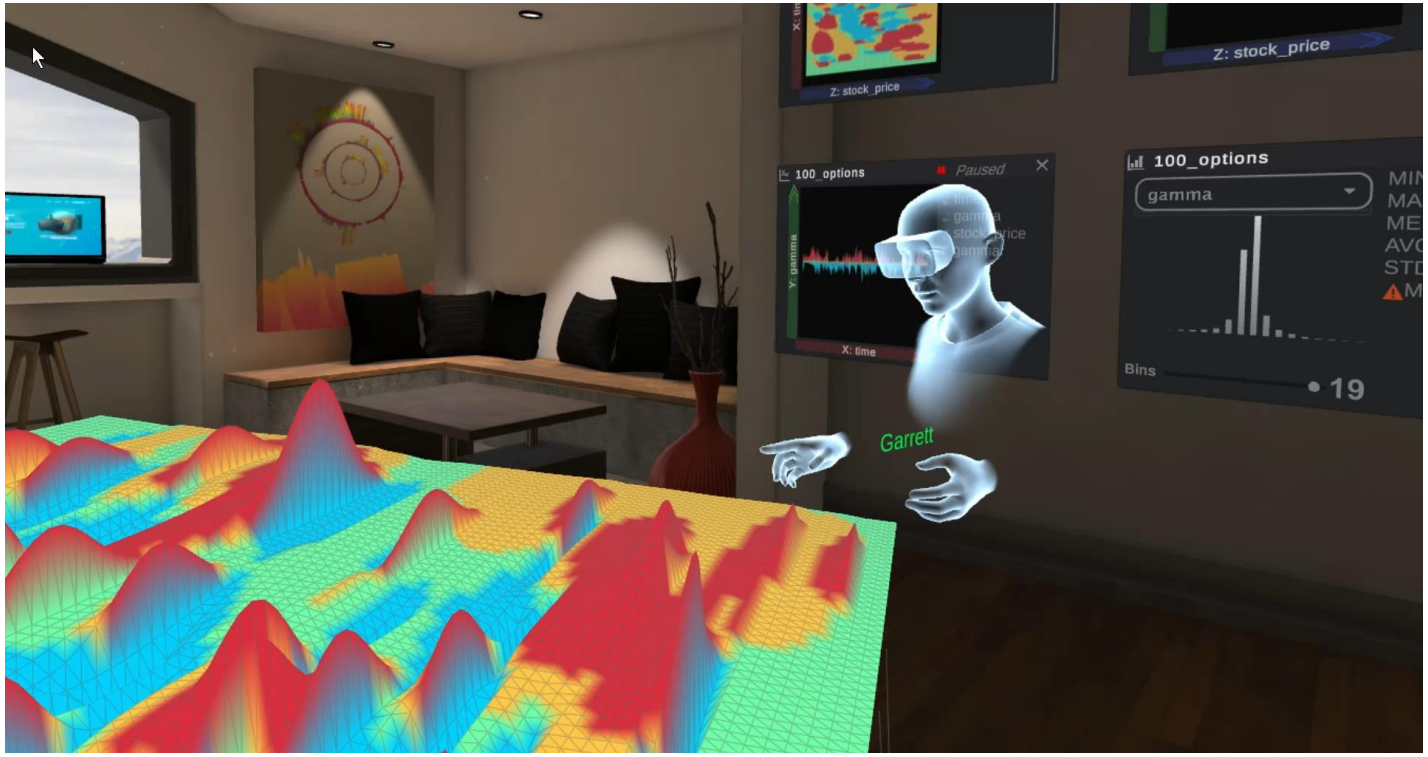
High-dimensional visualization

Clusters are clearly visible in 3D; on a “flat screen” still difficult to understand relative size and position…



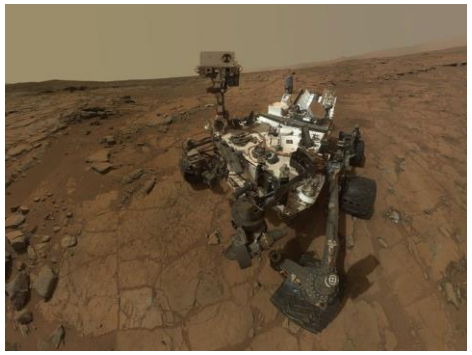
Immersive visualization

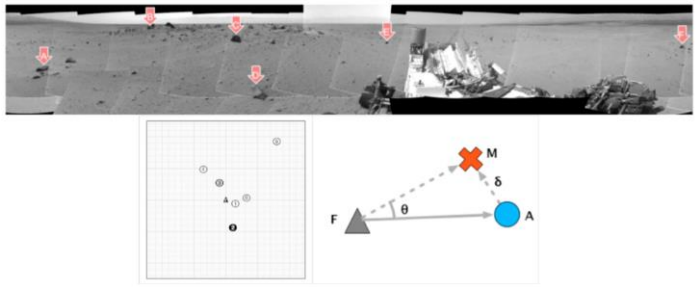
- Overcome limits of 3D on a “flat screen”
- Users can easily interact with their data and with each other in a familiar environment
- Easier and more effective collaboration across teams in different geographic locations
- Powerful tool for data analysis storytelling



Virtual environments offer new ways to visualize and interact with data

Case study: Mars map drawing - task





Problem: terrain on Earth is available for geologists to walk through. Scientists studying Mars cannot do that.

Goal: investigate whether the immersive environment provided scientists a more intuitive understanding of the terrain.

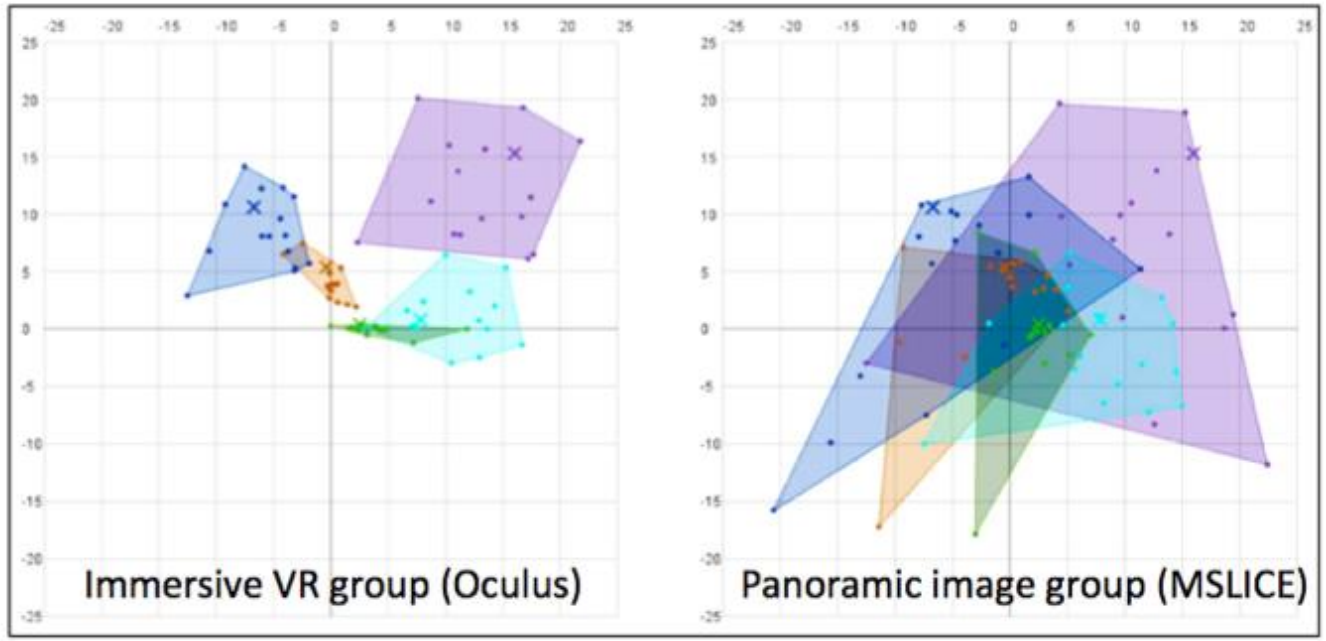
Experiment settings: map drawing task; six landmarks; scientists divided in 2 groups: one given a panorama image, one put in an immersive environment; compute angle and distance errors in map-making task (sum across all landmarks).

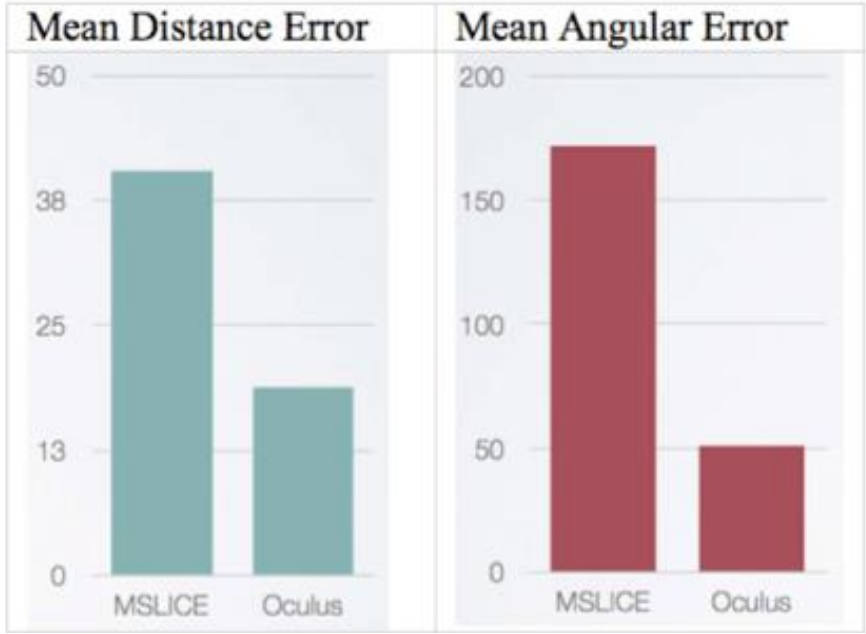
Experiments conducted at Caltech and NASA/JPL (Norris, Davidoff and Luo) showed that users commit less errors when immersed in a 3D environment.

Donalek C., et al., “Immersive and Collaborative Data Visualization Using Virtual Reality Platforms”, IEEE Big Data

Case study: Mars map drawing - results

Results: when relying on their natural perceptual capabilities in 3D, scientists are able to understand better their environment using immersive VR, as compared to a traditional, flat 2D image.





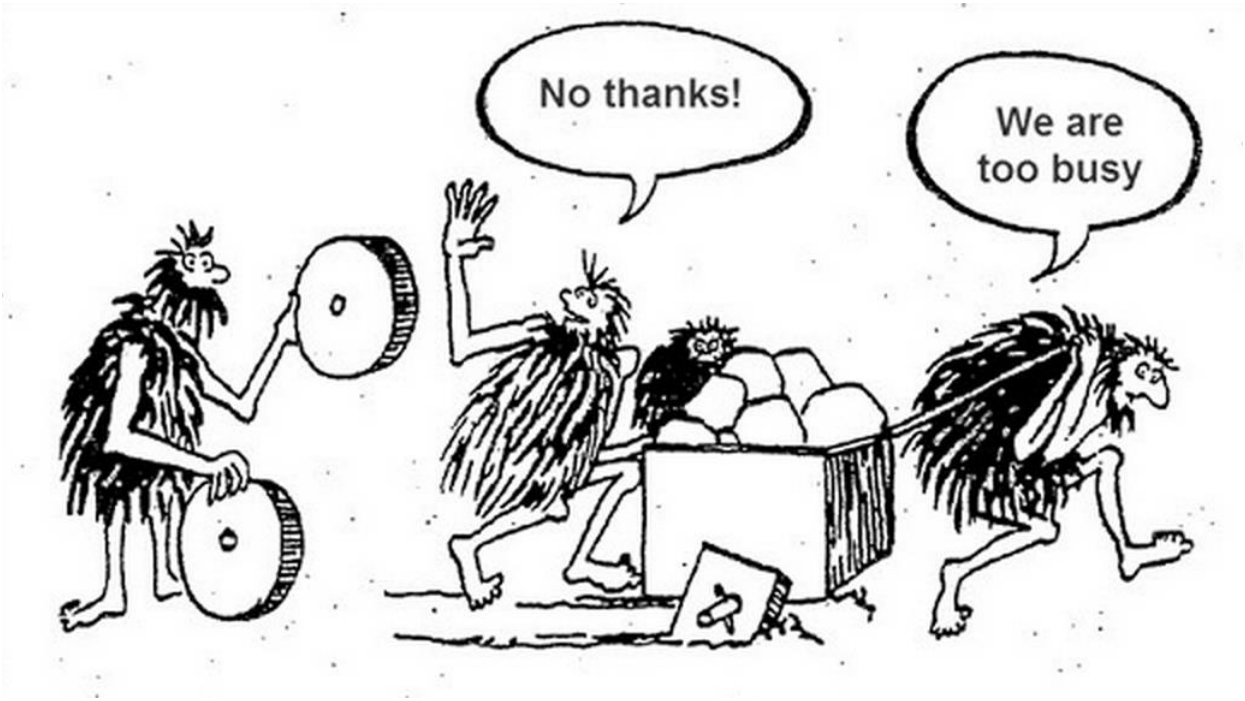
Points drawn by participants in the immersive VR group are more compact and closer to the correct location than the panorama group.

Understanding complex relationships

Ongoing study: Preliminary results

- Task: look at complex dataset with both traditional tools and VR
- Users in VR can understand relative size, position, more effectively
- Find answers more easily (e.g. outliers detection, trends)
- More effective collaboration

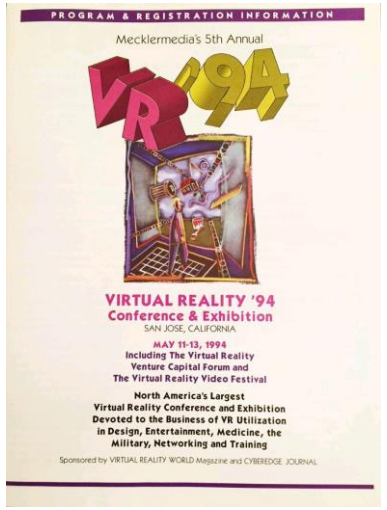
VR/AR: Skeptics and Natives



VR/AR: been around for 60 years…

Sensorama by Morton Helig (**1962**): first immersive, multisensory machine, still working today!

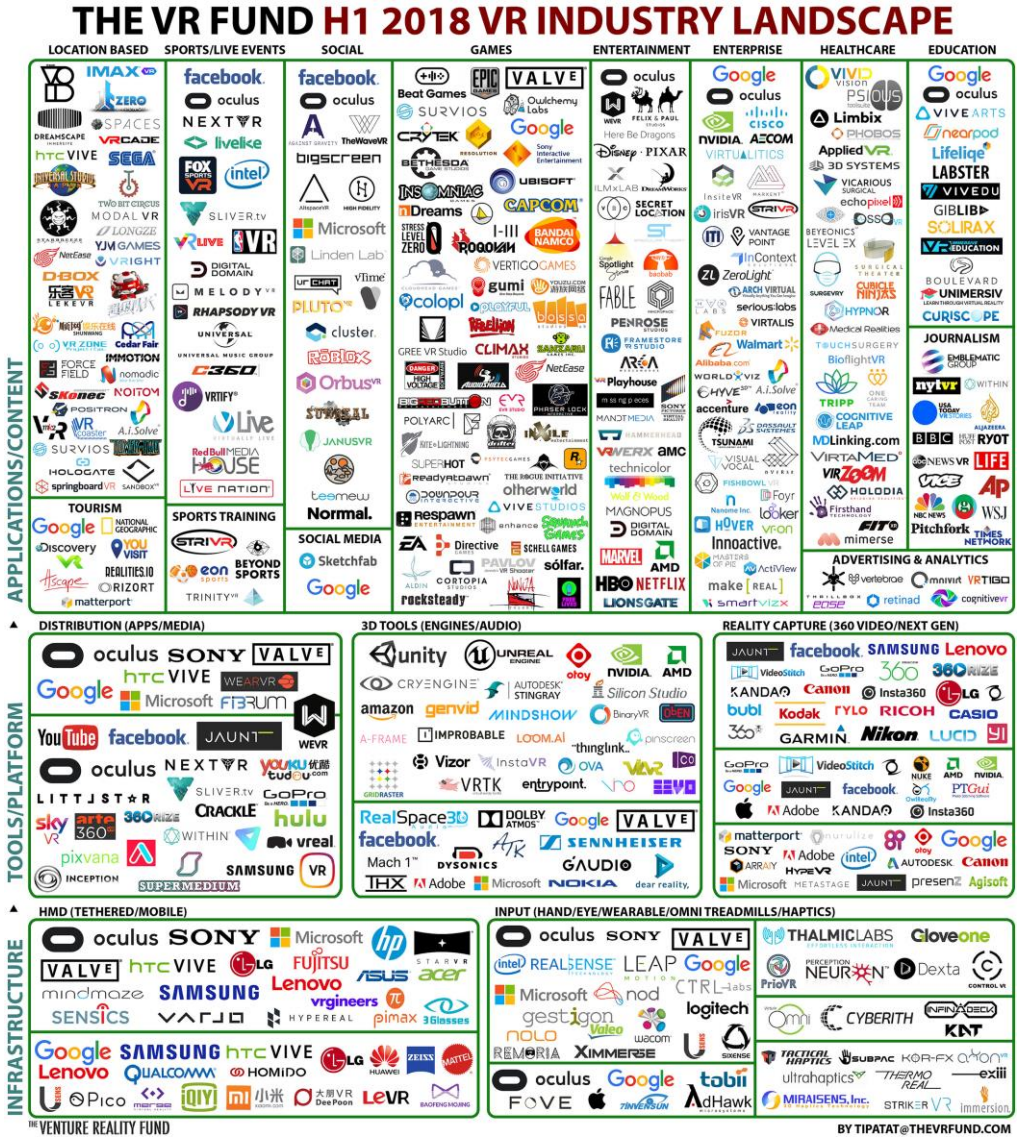
Ivan Sutherland & Bob Sproull (**1968**): the first augmented reality head-mounted display, suspended from a ceiling.

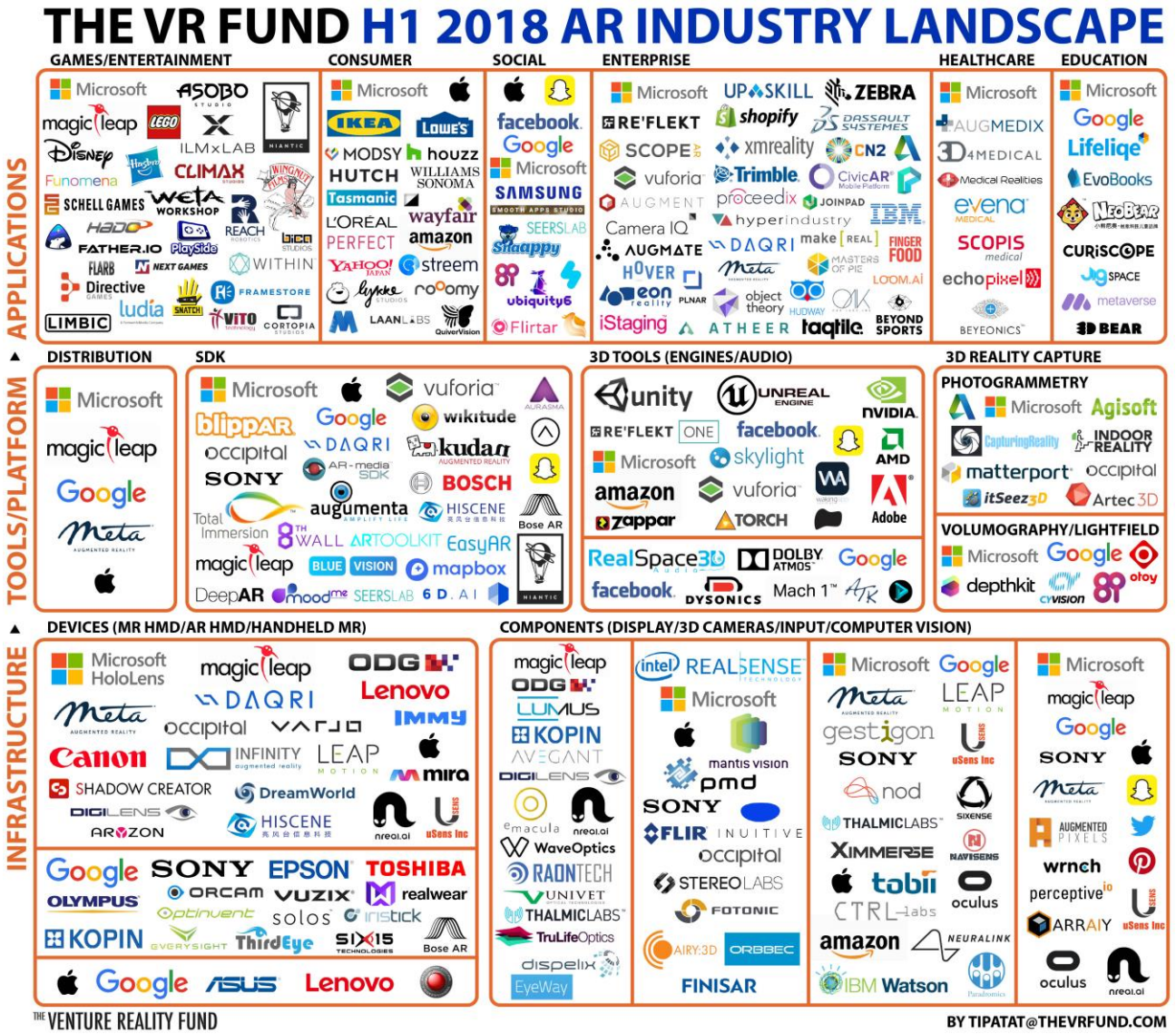


Virtual Boy Nintendo (**1995**). 32 bit, table-top, video game console. Advertises as the "*first portable console capable of displaying true 3D graphics*". Lasted 6 months on the market.

Nowadays: Major players involved

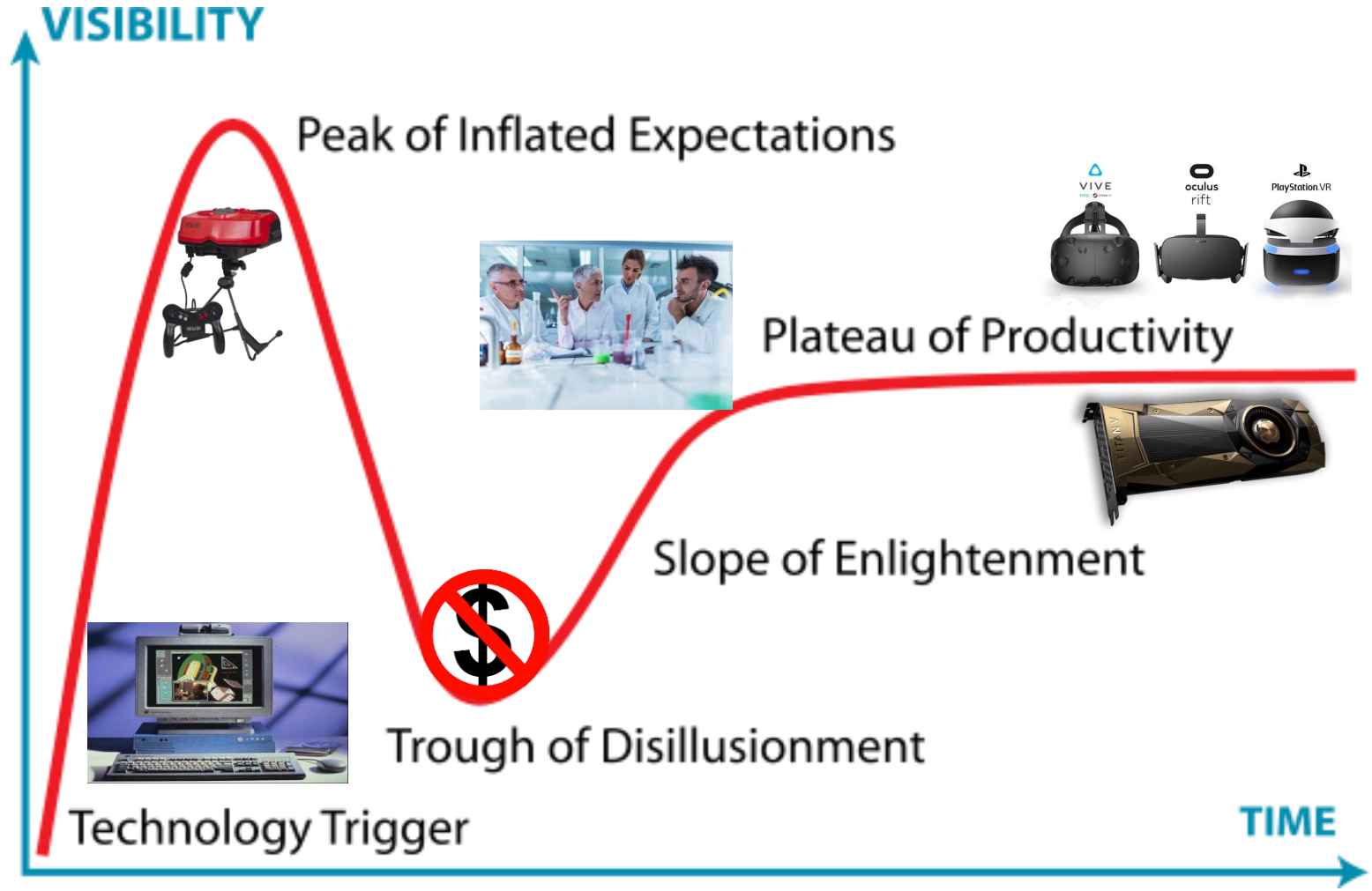
AR/VR industry landscape





VR is here to stay!

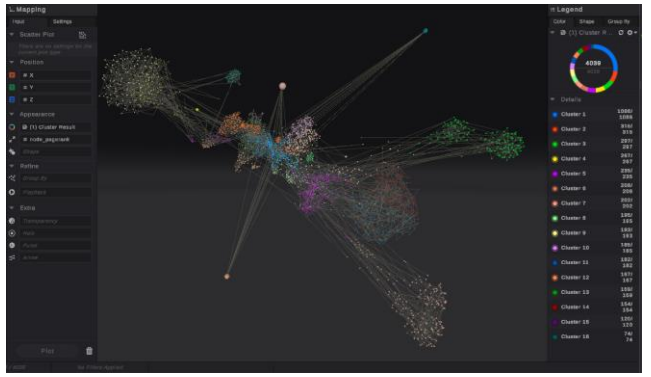
- Advances in computing and graphics: GPU era
- “By 2022, 70% of enterprises will be experimenting with immersive technologies for consumer and enterprise use, and 25% will have deployed to production” - Gartner report on 2019 technology trends

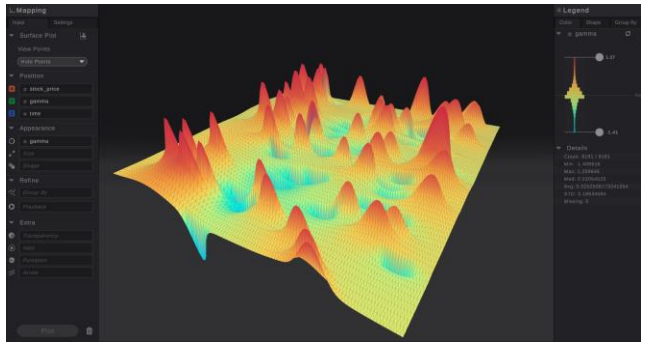


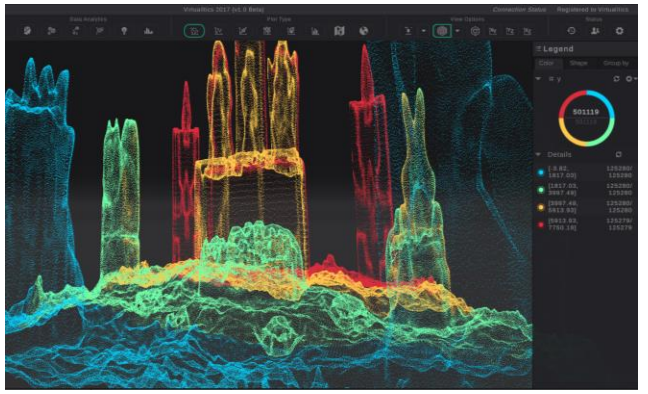
VR Hype Curve

GPU techniques in VIP

- Direct Compute: VIP uses Compute Shaders to optimize Force Atlas computation for Network Graph Visualizations
- Tessellation: VIP uses Tessellation Shader techniques to generate surfaces to visualize complex data in real-time
- Geometry shaders: Using point-topology meshes, VIP can render millions of data points while letting the GPU do the heavy lifting for creating complex 3D mesh detail for compelling visualization in VR







Research lead: Justin Gantenberg, Head of Engineering, Virtualitics

Case Study: large logistics company

KUEHNE+NAGEL

- Objective: need to make sense of massive supply chain data
- AI+VR to accelerate understanding of what processes are related to each other
- Collaborative environment
- Digital twin

Supply Chain Situation Room (concept developed in collaboration with Dr. Alban Quillaud – Head of Innovation and Analytics, Kuehne+Nagel)

Case Study: photogrammetry

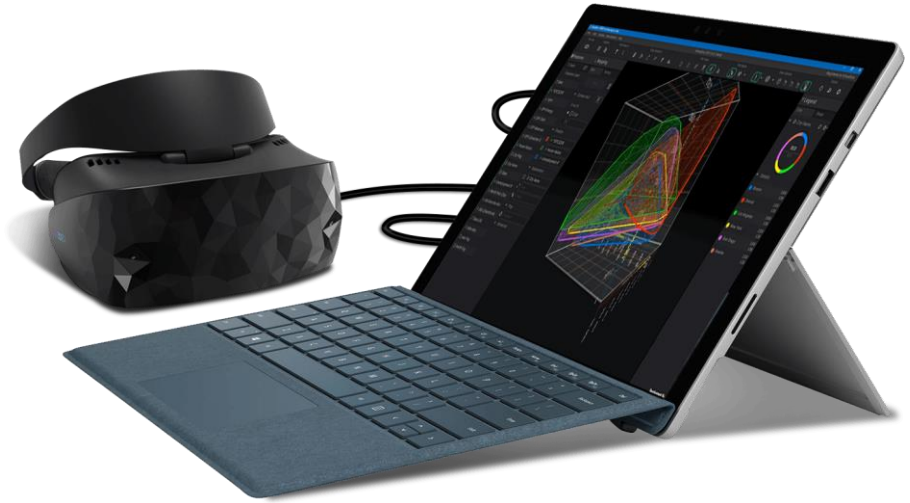
- High-end graphics cards are essential when developing content like photogrammetry
- Idea of endless screens to be immersed in data
- Vibrant colors to help aid users to focus on data
- Machine learning results superimposed on the map

Project lead: Greg Herboek, Head of Design, Virtualitics

So far so good, but…

What can slow down VR adoption

- Generational gap
- VR companies not delivering strong user experience
- Interaction
- Setup



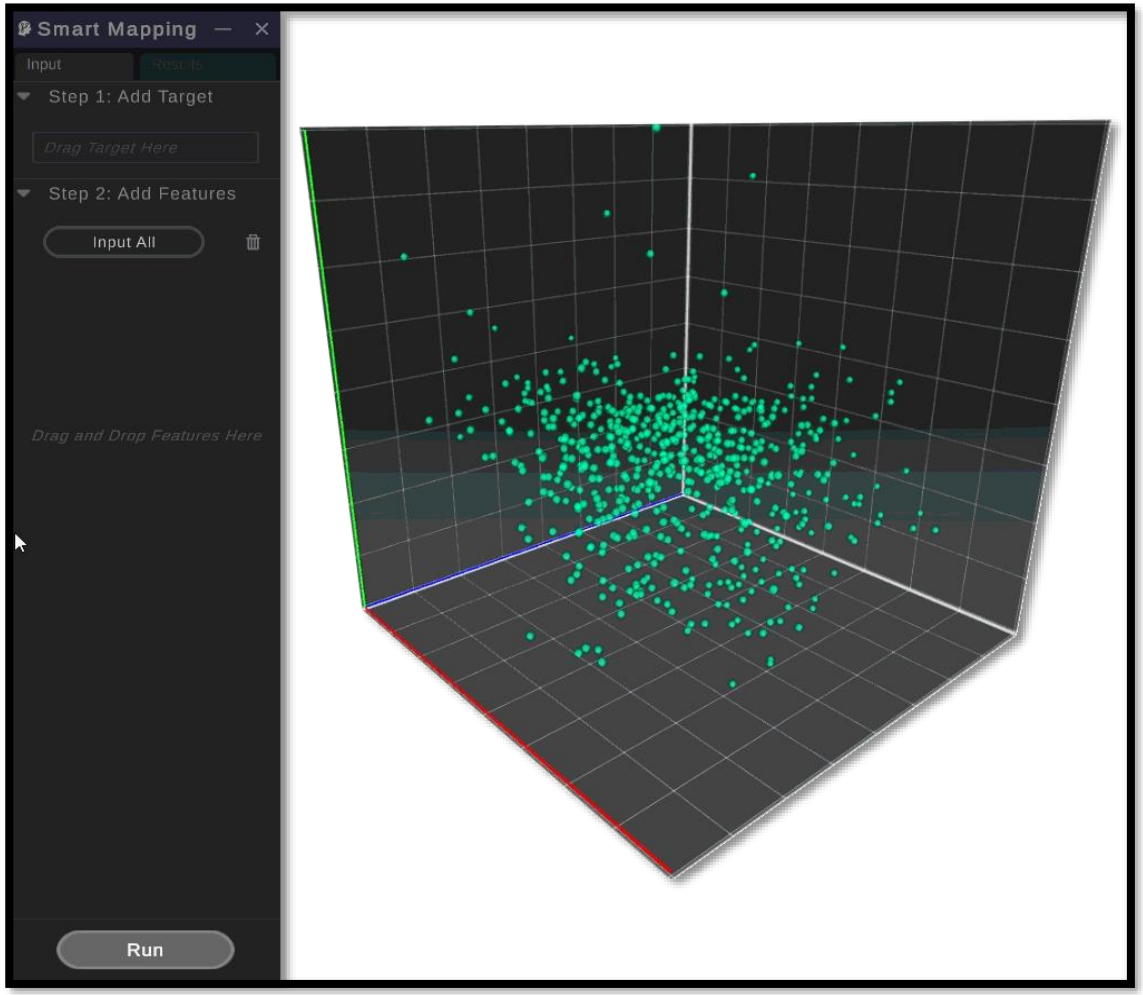
All-in-one devices and inside-out technology are making VR adoption easier

AI adoption: pain points

- Data scientists demand will increase ~30% by 2020 and there are not enough of them (source: IBM “The Quant Crunch” study)
- Businesses will look at “citizen data scientists” as a way to enable and scale data science capabilities: need AI-driven tools
- Difficult to convey the results to decision makers: AI seen as a black box

Solution: Democratizing machine learning

- Empower non-data scientists through easy-to-use, embedded ML routines
- Provide insights in natural language
- AI+VR makes data insights more broadly shareable across the business, including analysts and decision makers



Smart Mapping: proprietary Virtualitics routine, drag-and-drop functionality to find key insights using AI and high-dimensional visualization

How to explain complex ML models

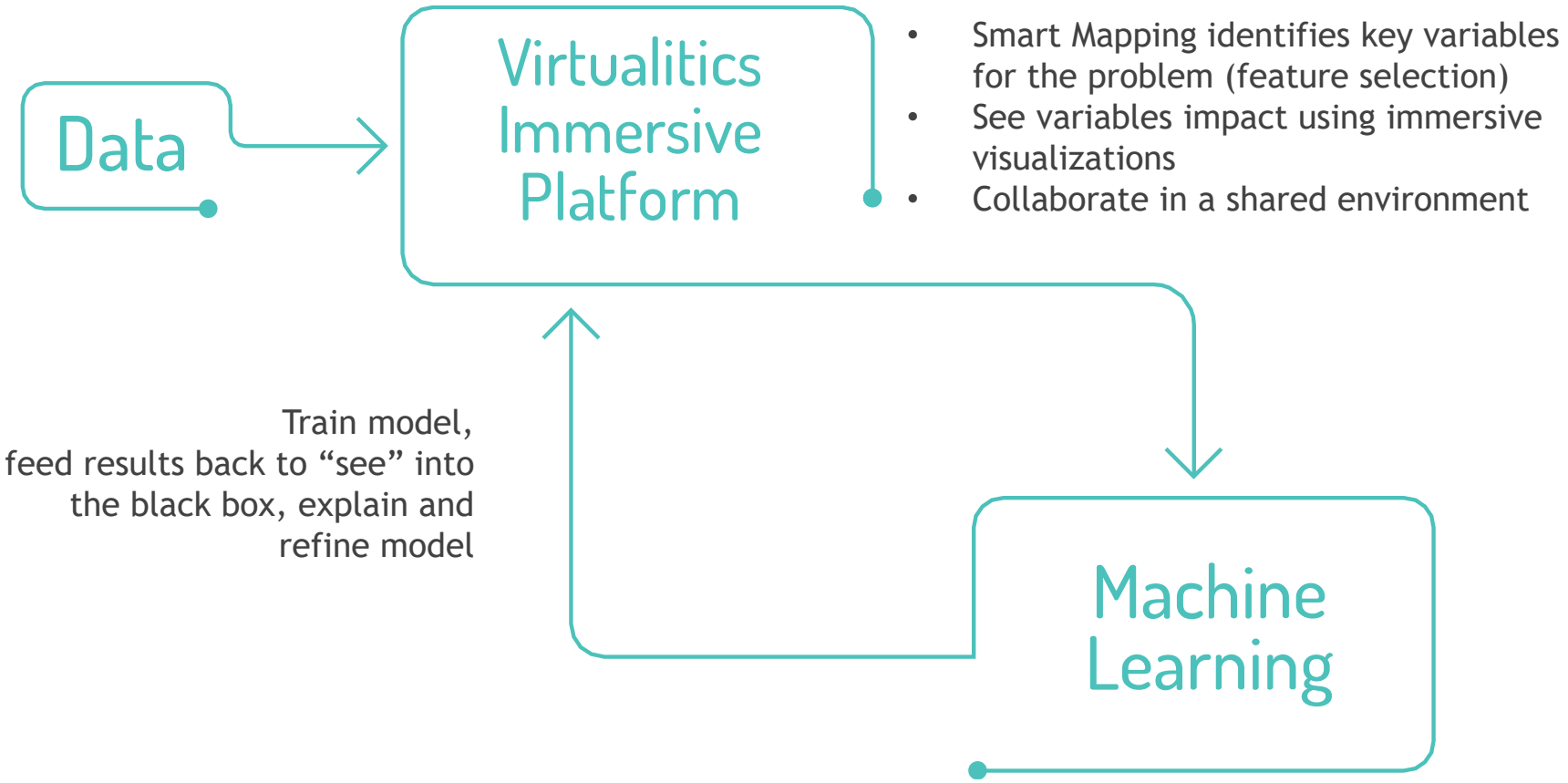
Case Study: "Step into the box"

AI-explainability: Use immersive visualization to describe complex Machine Learning models.

Convey information to upper management, bridge the gap between data scientists and decision makers.

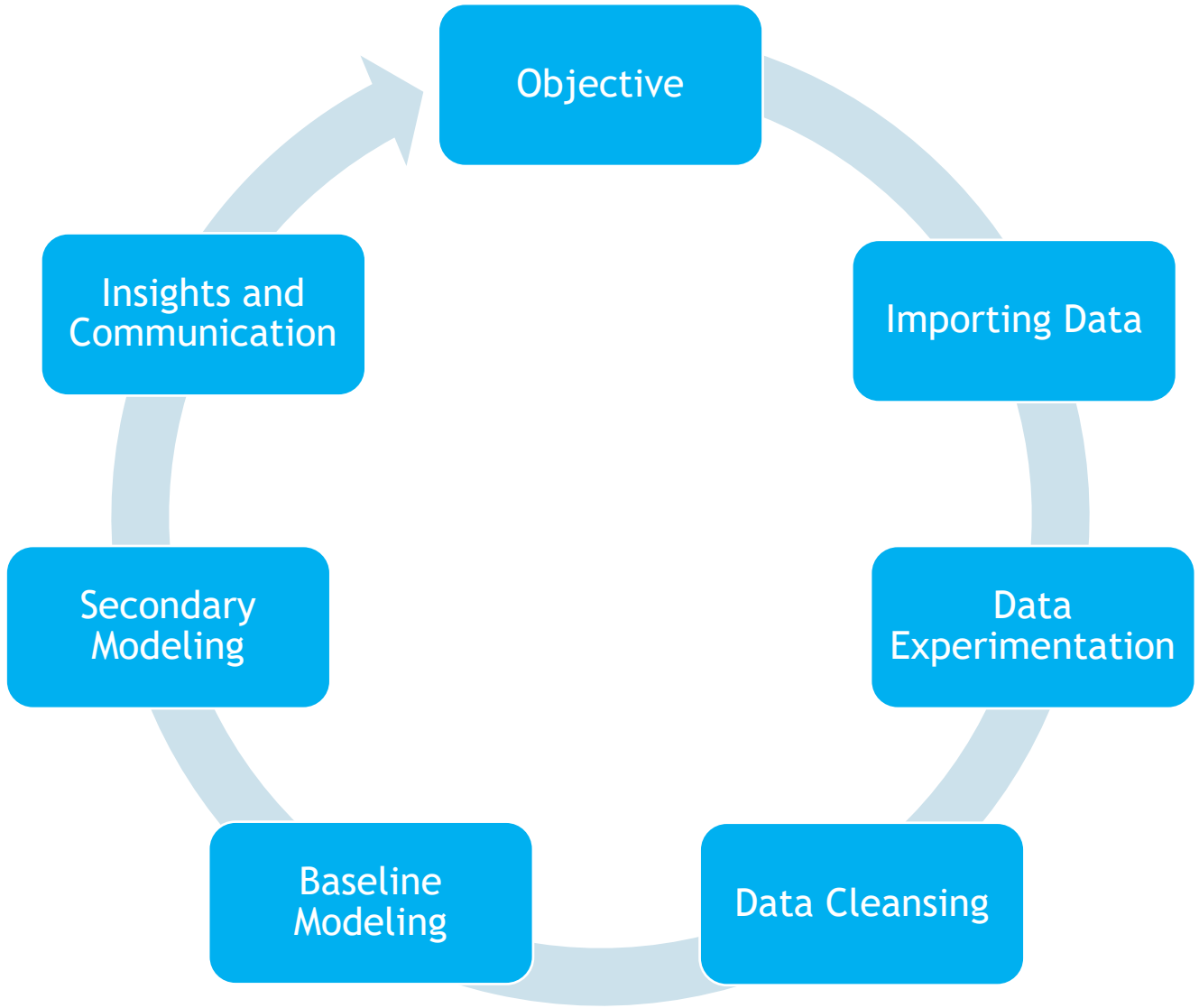
Case study: large financial company

AI + VR pipeline: Prediction and Visualization + Collaboration between teams across the country



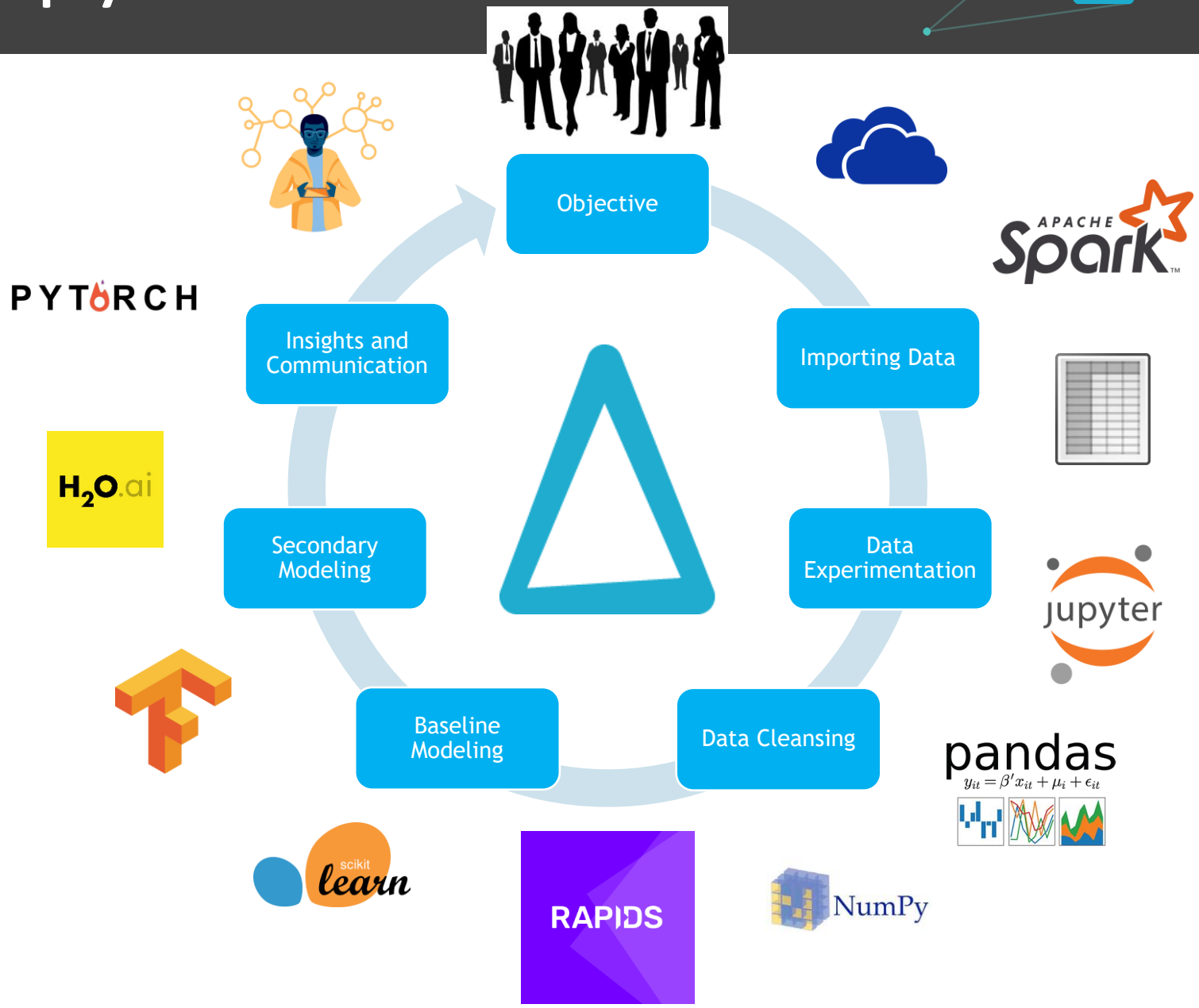
Modern data science

- Involve complex tech stacks with moving parts
- Highly iterative – optimize one step at a time
- Need to be explained to business stakeholders



Virtualitics API - pyVIP

- Designed for use with industry leadings tools
- Immersive and interactive way to analyze results from leading packages and proprietary tools
- Faster and improved iterations in data science workflow
- Explainable results on-the-fly ready for collaboration



JupyterLab | Explore Finance - 1

localhost:8888/lab

File Edit View Run Kernel Tabs Settings Help

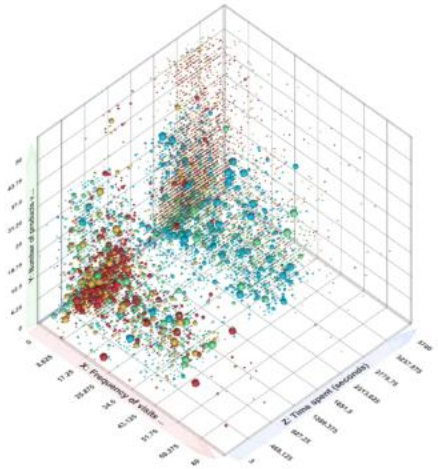
Exploring ecommerce dat | Testing.ipynb | Exploring Geospatial Data

Code | Python 3

Smart Mapping - key drivers of 'Amount Spent'

[44]:
```
vip.smart_mapping(data["Amount spent (USD)"], exclude=["User ID"])
```

	SmartMapping Rank	Feature	Correlated Group
0	1	Sources	None
1	2	Married	None
2	3	Frequency of visits last week	None
3	4	Number of products viewed	None
4	5	Time spent (seconds)	None



VIP's AI based Insights

Specifically lets find insights about our top quartile of spenders!

[ ]:
```
data['SlowBulkShoppers'] = vip.get_visible_points()
```

[ ]:
```
data["ExpressShoppers"] = vip.get_visible_points()
```

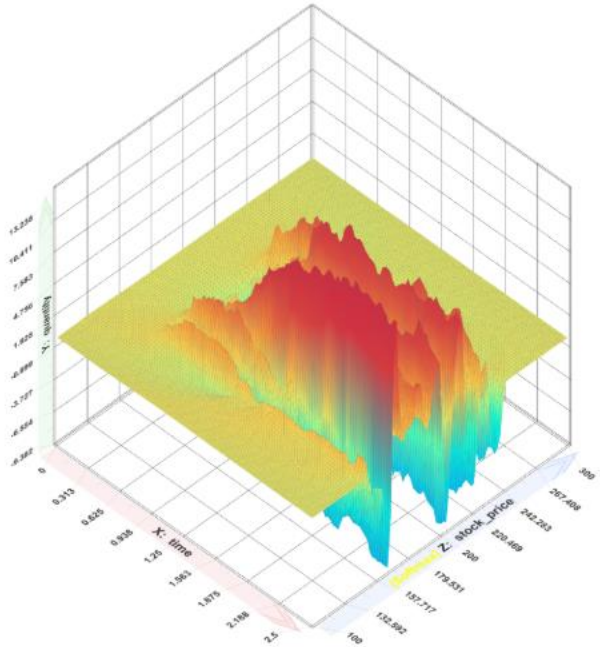
Exploring ecommerce data.i | Untitled.ipynb

Code | Python 3

```
Connection Successful! Initializing session.
```

[4]:
```
data = pd.read_csv("D:/Demo Datasets/gamma (1).csv")
vip.load_data(data, "Trading Risk")
```

[5]:
```
vip.surface(x=" time", y=" quantity", z=" stock_price", color=" quantity",
            color_type="grad", color_normalization="Softmax", z_normalization="Softmax")
```



Influence from the energy sector

[6]:
```
data = pd.read_csv("D:/Demo Datasets/commodities.csv")
vip.load_data(data, "Energy Sector")
```

[7]:
```
vip.line(color="series_name", x="Date", y="Value", y_normalization="log10", x_scale=2.0)
```

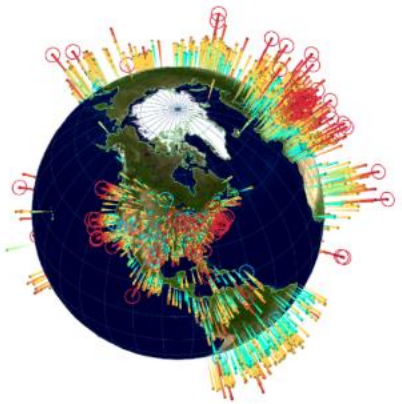
VB-384 Ellipsoid Plot.ipynb | Exploring Twitter GOT_v3.ipy

Code | Python 3

Now lets look for tweets that are anomalous in terms of sentiment!

[31]:
```
anoms = vip.threshold_ad(threshold=1, features=["Sentiment"])
```

```
Threshold Anomaly Detection completed with a threshold of 1%.  Anomalies ar
e marked in the plot with Halos.
```



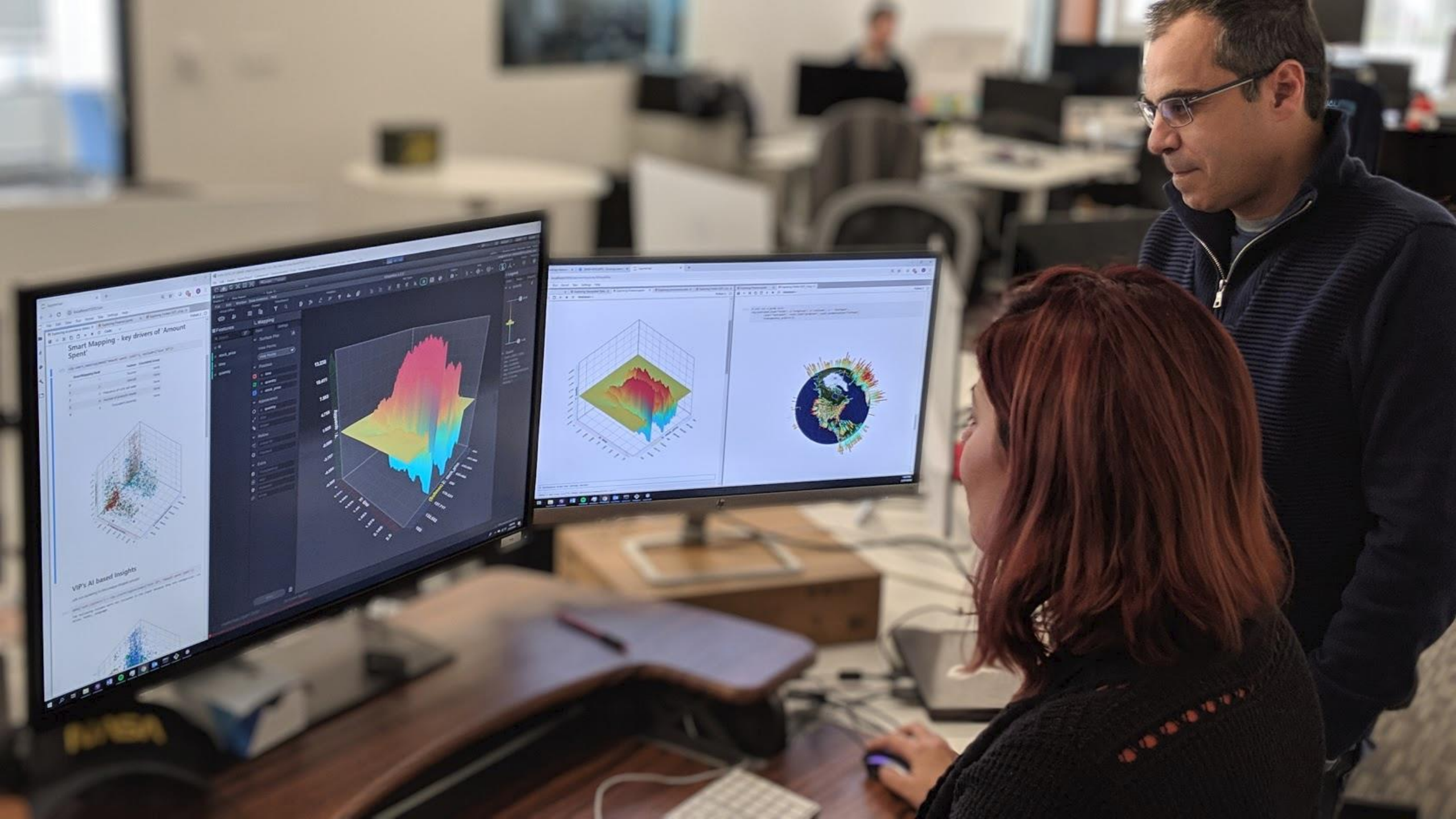
Using TF-IDF to Vectorize the tweet text

We are going to turn the tweet text into a vector by using Tfidf vectorizer with a certain vocabulary. This will allow us to visualize the content of the tweets. First we will start with a generic vocabulary of 100 words with a lot of meaning.

[24]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
import numpy as np

vocab = pd.read_csv('GOT-Twitter/mostUsefulWords.csv').values
# The csv stores words using a numbering scheme like this:
#    1. word1
#    2. word2
#
```

5:13 PM
2/24/2019

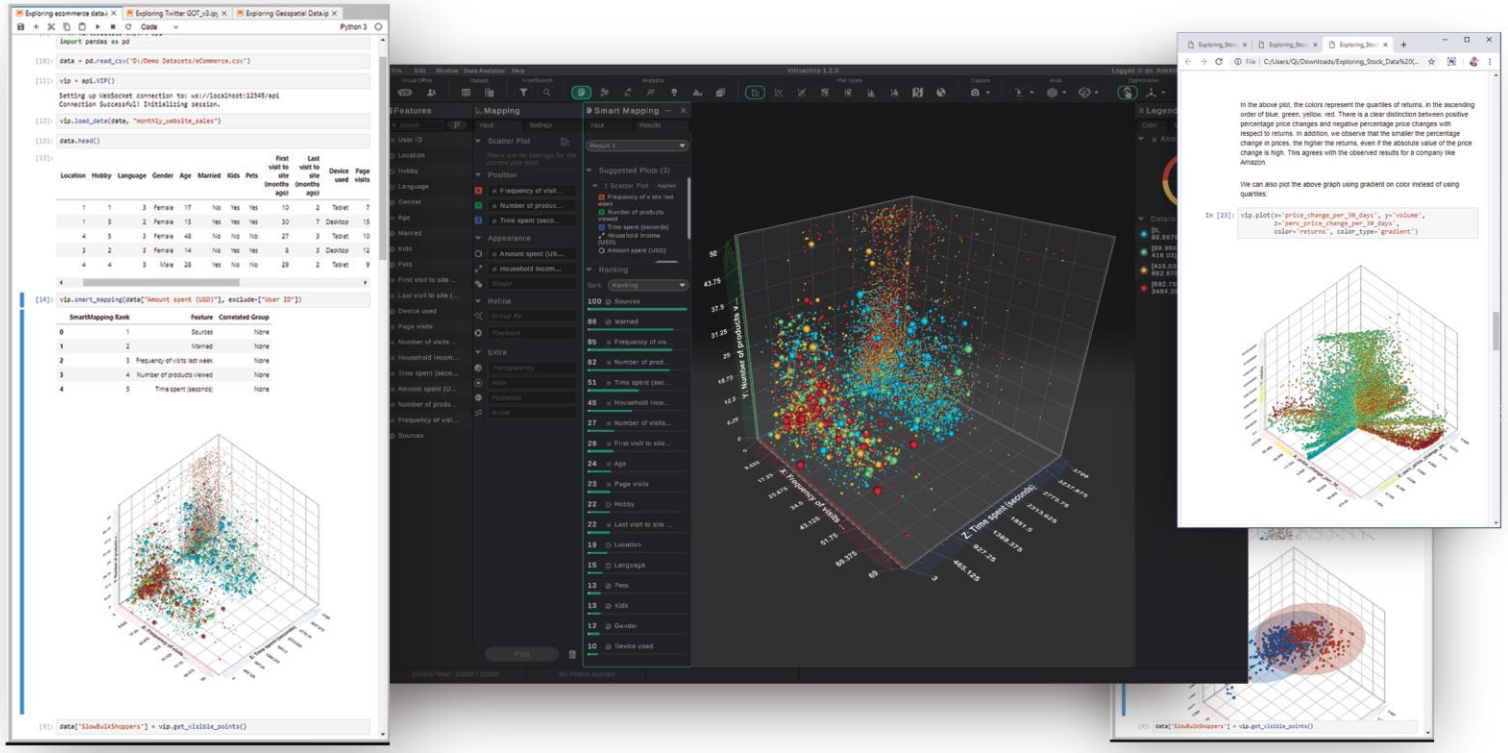

Smart Mapping - key drivers of 'Amount Spent'
VIP's AI based insights

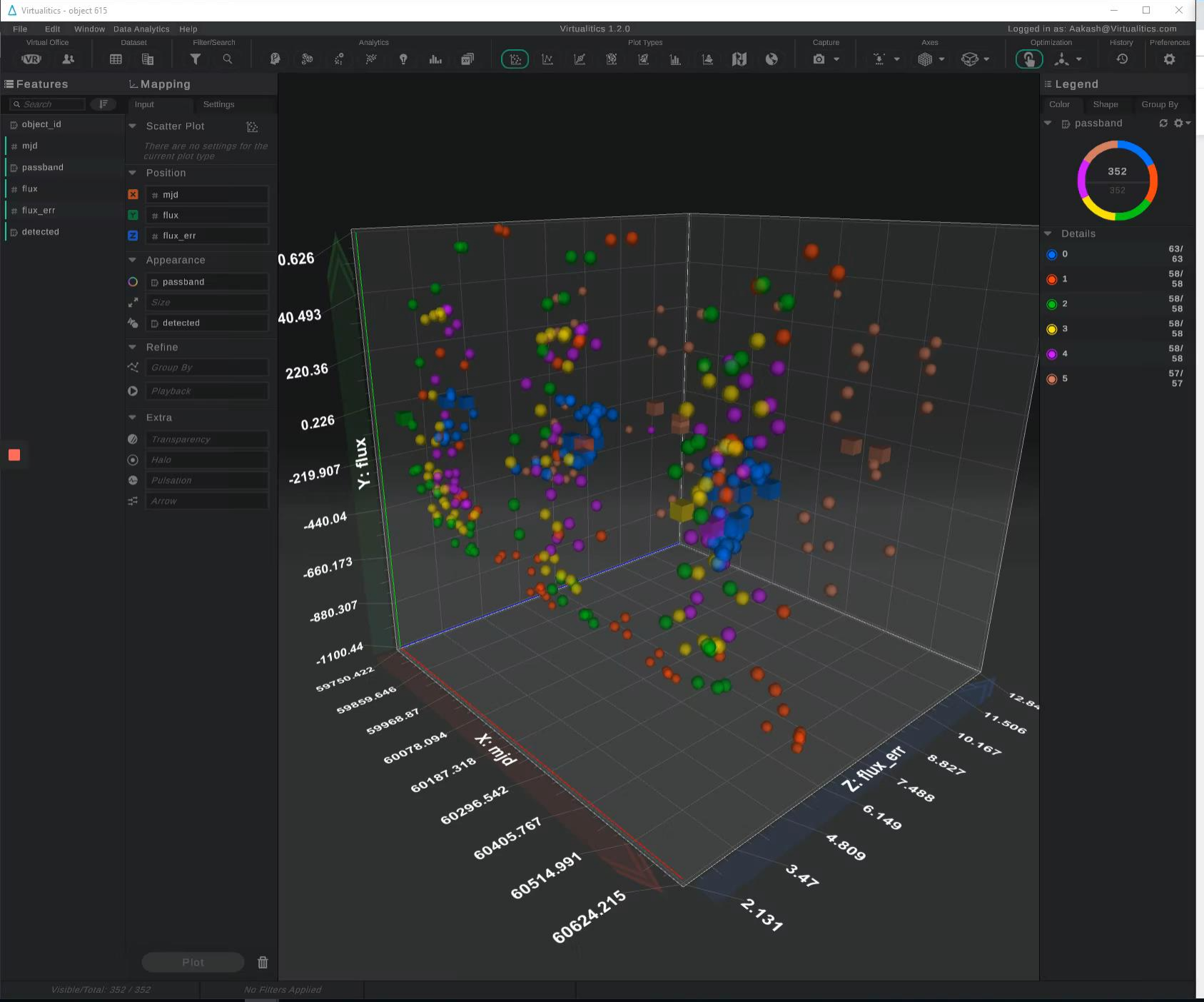
Connect VIP with NVIDIA HPC Clusters

HPC Cloud Cluster

1. Launch Jupyter on HPCC
2. Import pyVIP and specify host/port
3. Setup SSH Tunneling

Virtualitics API uses WebSockets to connect with VIP. The Virtualitics API can be used on any Python ready platform!





ngrok - secu | ngrok - Stat | Compute Er | Home | rapids_lsst_g

localhost:8890/notebooks/rapids_lsst_gpu_only_demo.ipynb#Data-Exploration

jupyter rapids_lsst_gpu_only_demo (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Markdown

```
# test_flux_skew_gd = test_gd
train_flux_skew_gd = train_gd
```

```
CPU times: user 129 ms, sys: 4.13 ms, total: 133 ms
Wall time: 132 ms
```

In [ ]:

Load data for the ETL part 2 with 11x speedup

In [54]:
```
%%time
# read data on gpu
step = 'load data part2'
start = time.time()
ts_cols = ['object_id', 'mjd', 'passband', 'flux', 'flux_err', 'detected']
ts_dtypes = ['int32', 'float32', 'int32', 'float32','float32','int32']

# test_gd = gd.read_csv('%s/test_set.csv'%PATH,
#             names=ts_cols,dtype=ts_dtypes,skiprows=1+SKIP_ROWS) # skip the hea
train_gd = gd.read_csv('%s/training_set.csv'%PATH,
            names=ts_cols,dtype=ts_dtypes,skiprows=1)

cols = ['object_id', 'ra', 'decl', 'gal_l', 'gal_b', 'ddf',
       'hostgal_specz', 'hostgal_photoz', 'hostgal_photoz_err',
       'distmod','mwebv', 'target']
dtypes = ['int32']+['float32']*4+['int32']+['float32']*5+['int32']

train_meta_gd = gd.read_csv('%s/training_set_metadata.csv'%PATH,
            names=cols,dtype=dtypes,skiprows=1)
# del cols[-1],dtypes[-1]
# test_meta_gd = gd.read_csv('%s/test_set_metadata.csv'%PATH,
#             names=cols,dtype=dtypes,skiprows=1)
GPU_RUN_TIME[step] = time.time() - start
```

```
CPU times: user 86.9 ms, sys: 28.3 ms, total: 115 ms
Wall time: 114 ms
```

RAPIDS ETL 100X Speed Up

In [53]:
```
%%time
# GPU
start = time.time()
step = 'ETL part2'
train_final_gd = etl_gpu(train_gd,train_meta_gd)
train_final_gd = train_final_gd.merge(train_flux_skew_gd,on=['object_id'],how='l
GPU_RUN_TIME[step] = time.time() - start
```

```
CPU times: user 298 ms, sys: 84.4 ms, total: 383 ms
Wall time: 381 ms
```

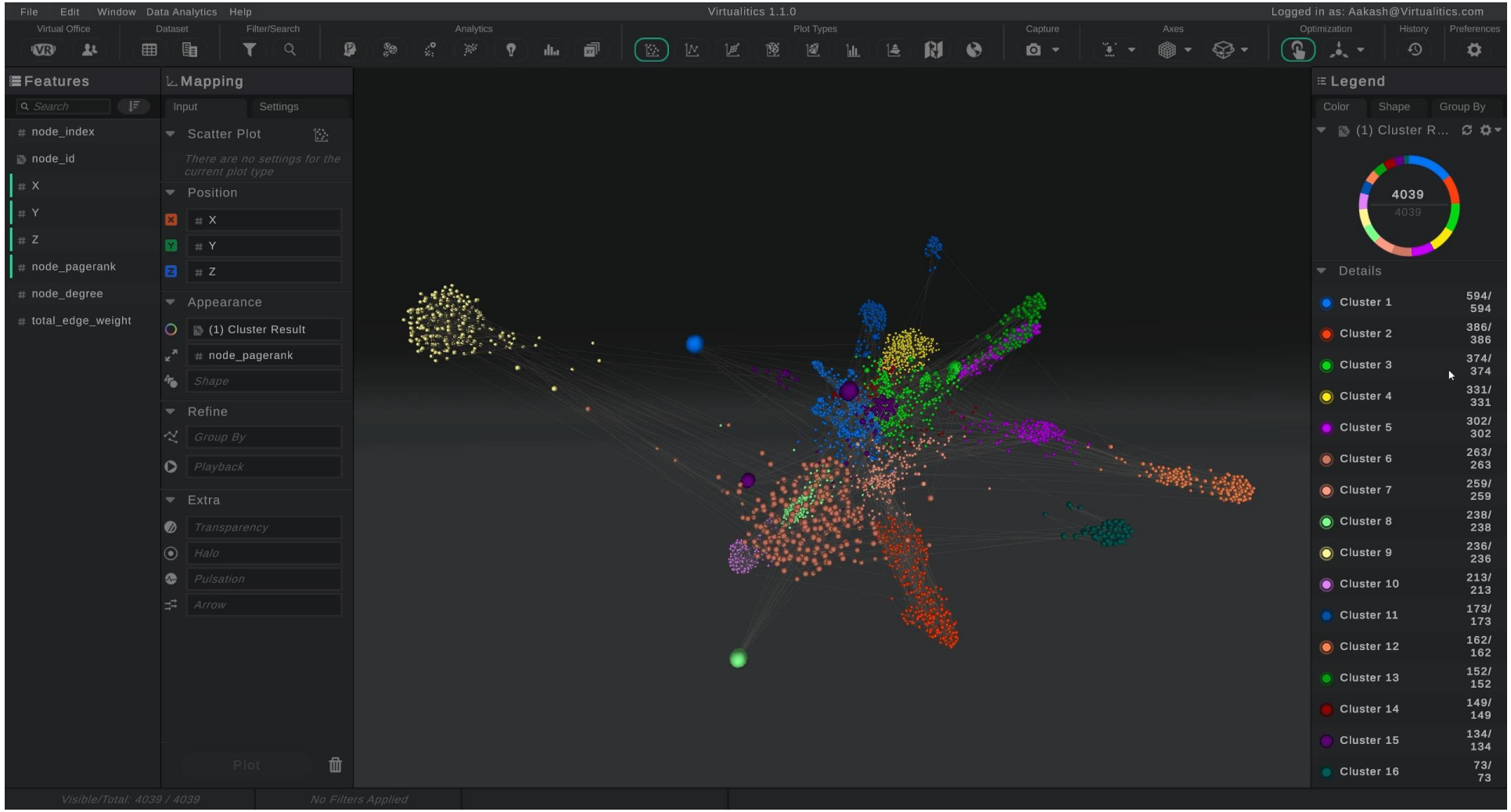
Data Exploration

In [23]:
```
train = train_final_gd.to_pandas()
vip.load_data(train, "training_data")
vip.scatter(x="hostgal_specz", y="hostgal_photoz", z="skew_flux",
            shape="mean_passband", color="target")
vip.normalize("Softmax")
```

In [27]:
```
vip.pca(num_components=5, exclude=["object_id"])
vip.normalize("Softmax")
```

1:49 PM 3/7/2019

Network graphs



Join us for a demo!

- AI+VR = Augmented Analytics
- AI can guide users through their data and unlock key insights
- VR provide immersive visualization of insights and also contribute to unique collaborative environment

DESKTOP+VR

AI+VR

DISCOVER

SHARE