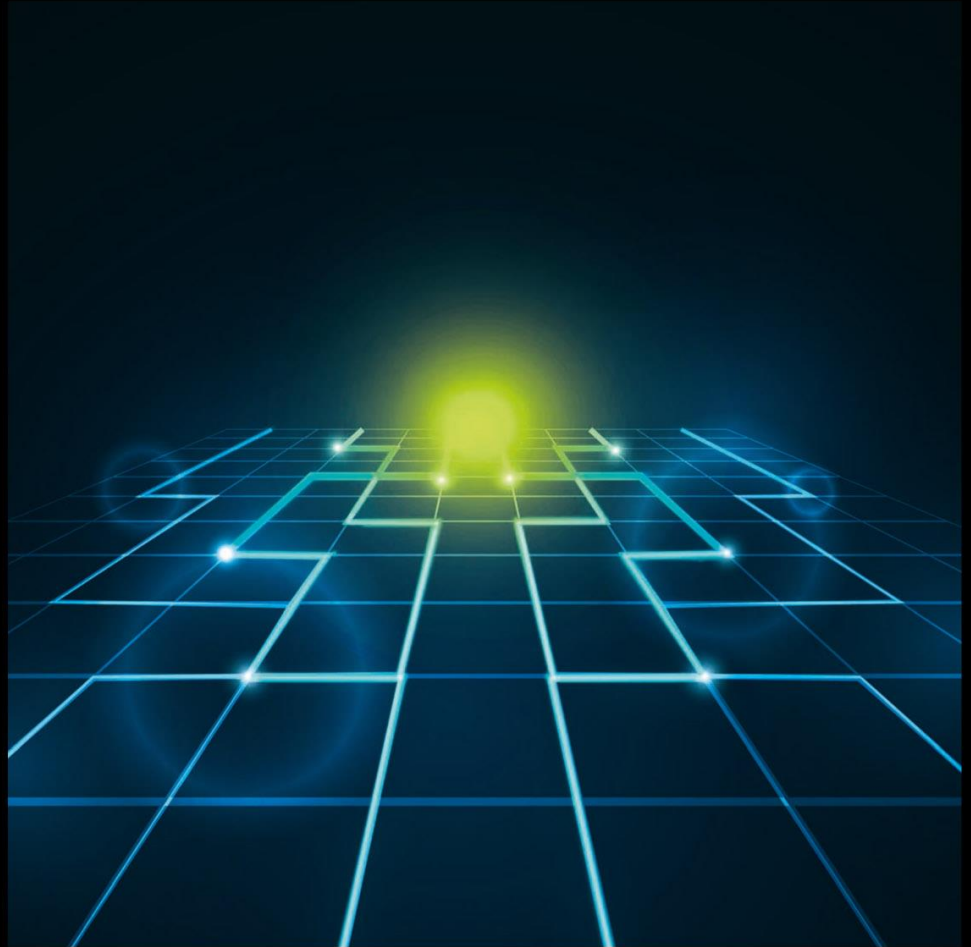


GI Glimpse



**Actuaries
Institute**

3 August 2016 • Sydney





Software and Tools

Hugh Miller

Taylor Fry

© Taylor Fry

*This presentation has been prepared for the Actuaries Institute 2016 GI Glimpse Seminar.
The Institute Council wishes it to be understood that opinions put forward herein are not necessarily those of the
Institute and the Council is not responsible for those opinions.*



Software and Tools

Definition:
*A momentary
or cursory view*

Hugh Miller

Taylor Fry

© Taylor Fry

*This presentation has been prepared for the Actuaries Institute 2016 GI Glimpse Seminar.
The Institute Council wishes it to be understood that opinions put forward herein are not necessarily those of the
Institute and the Council is not responsible for those opinions.*



Why software and tools?

- What a lot of us do a lot of the time
- Fundamentally determines what you can do (and how much of it!)



What's our target application today?

“Actuarial analysis to support GI”

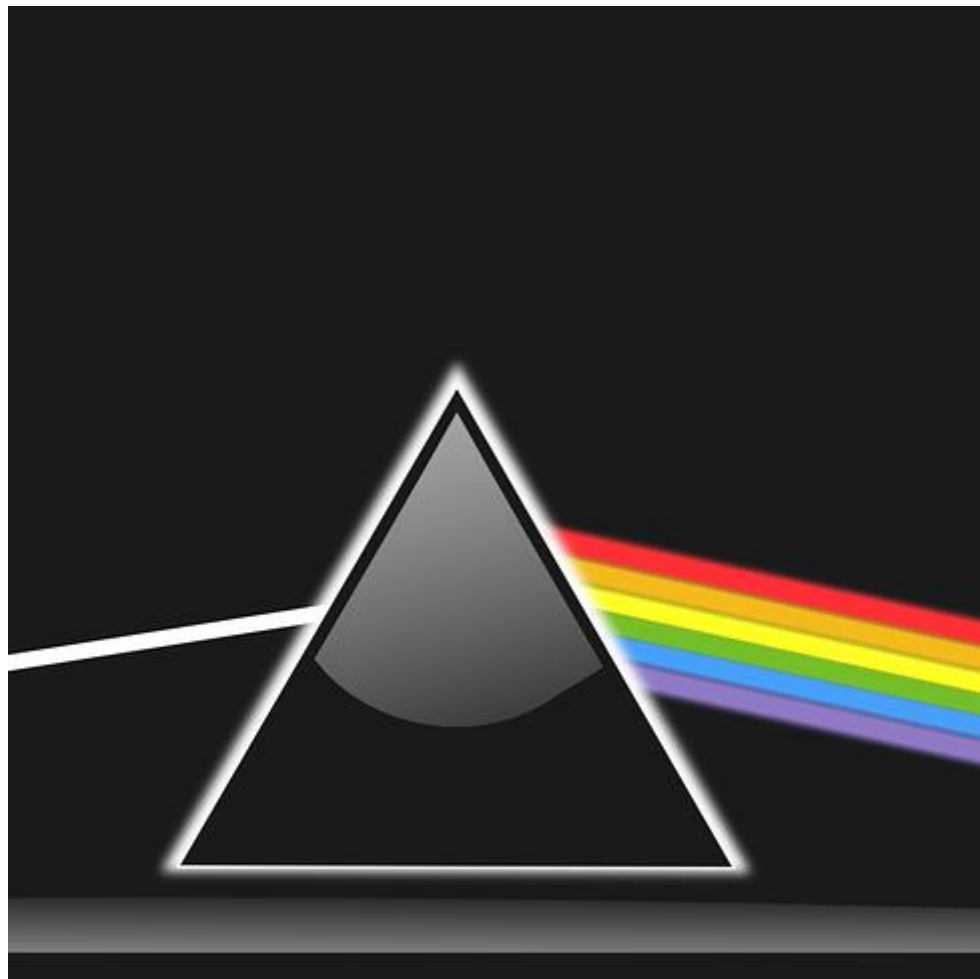
- Pricing models
- Other customer models
- Statistical case estimation
- Monitoring

This talk overlaps with:

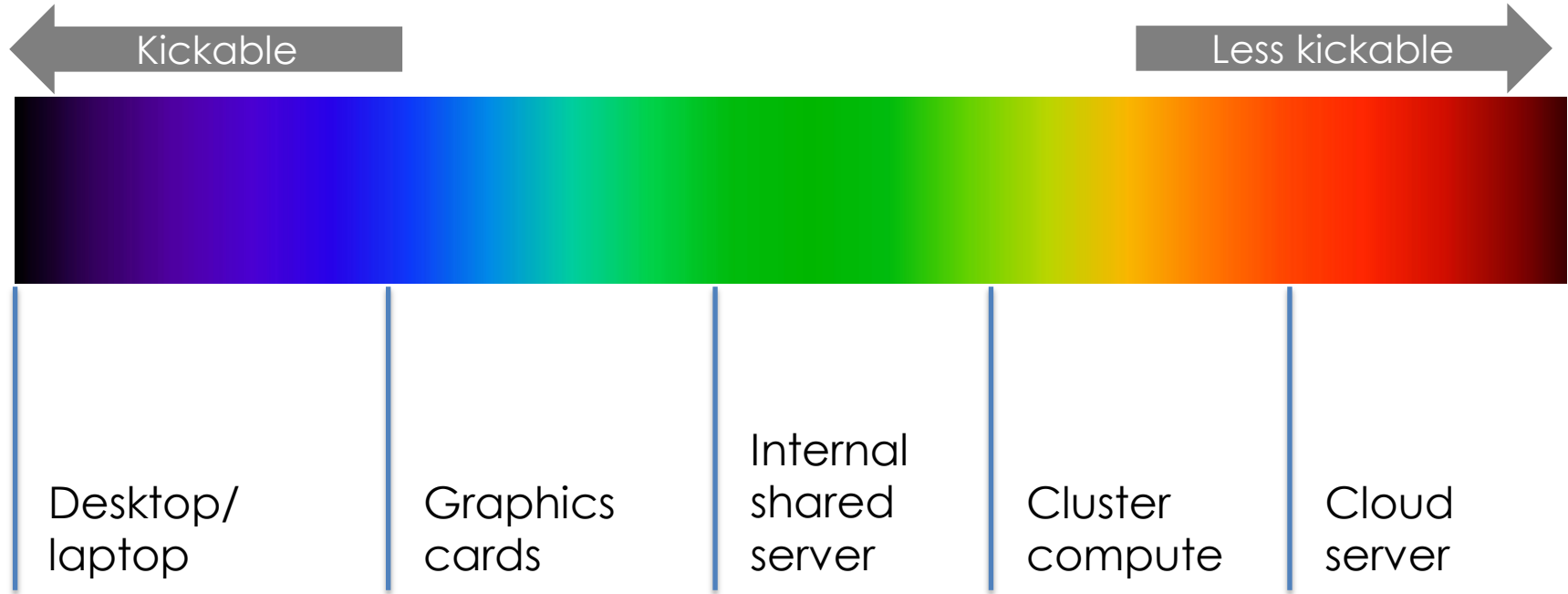
- *Software choices for actuaries.* Hugh Miller, GI Seminar, 2014
- *Data & Software.* Hugh Miller, Data analytics seminar, 2015.

A set of spectrums

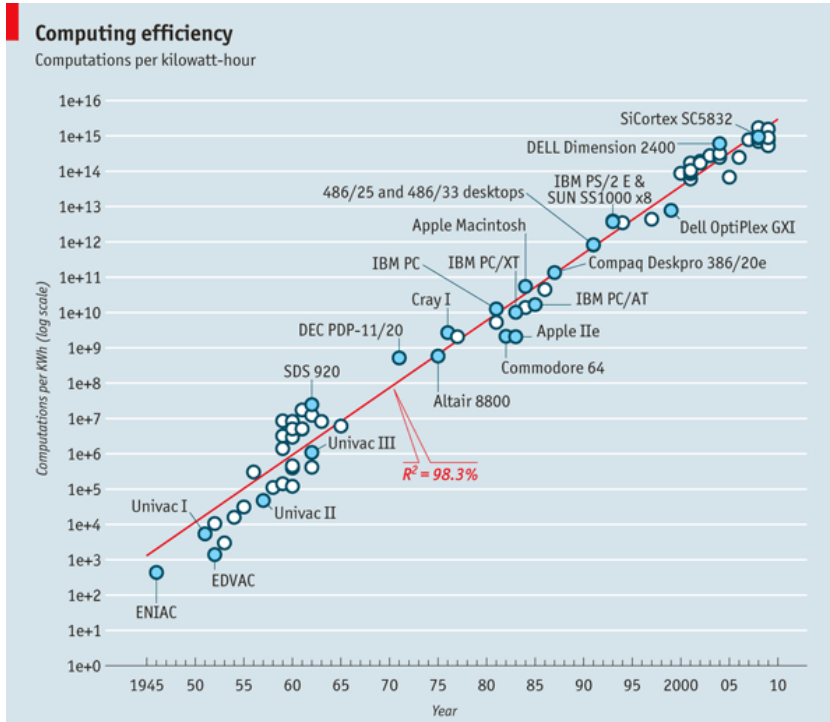
- Hardware choice
- Software choice
- Algorithm choice
- Communication choice



Hardware choice

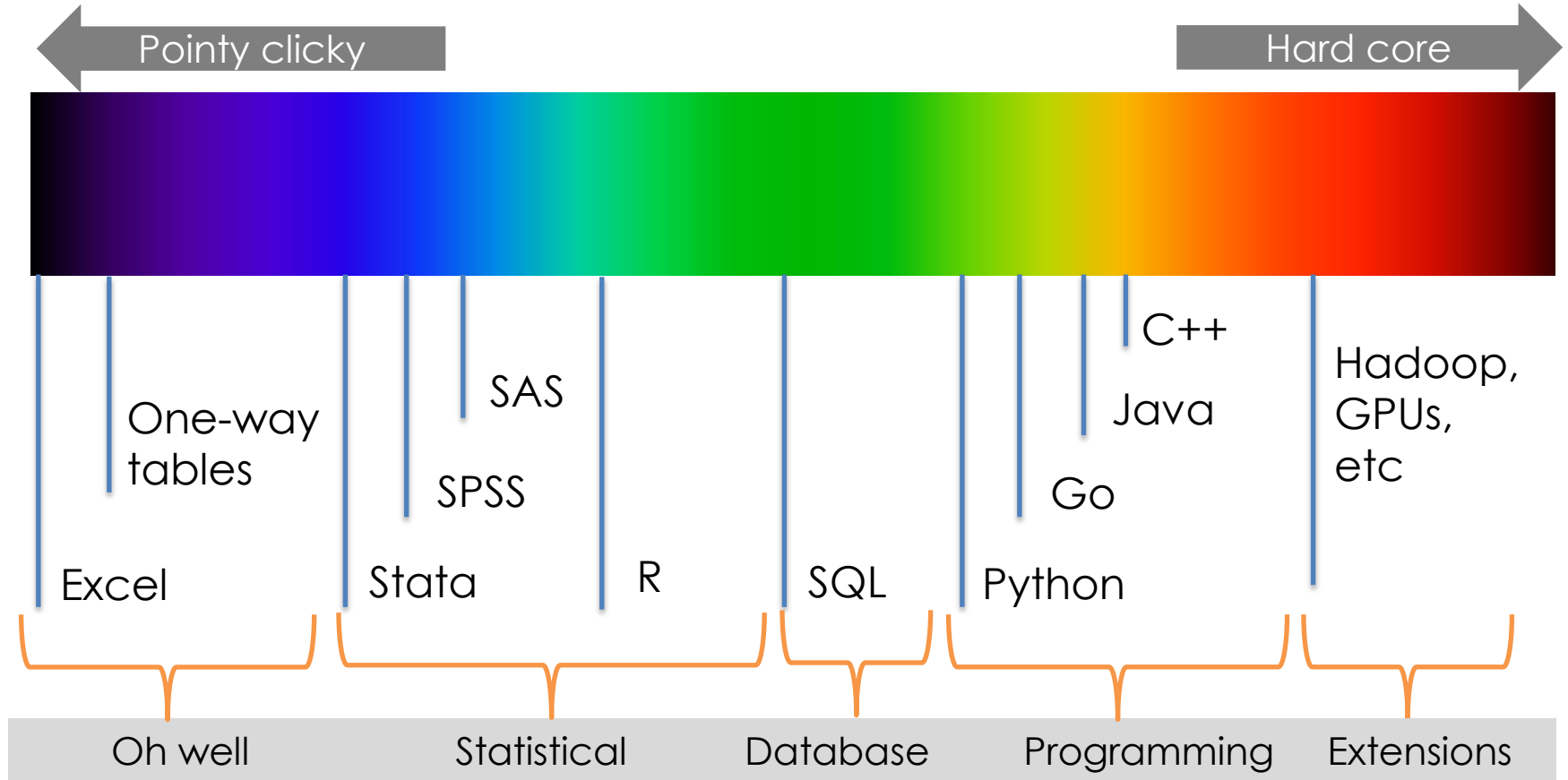


Hardware will continue to improve



Source: Jonathan Koomey via the Economist

Software choice

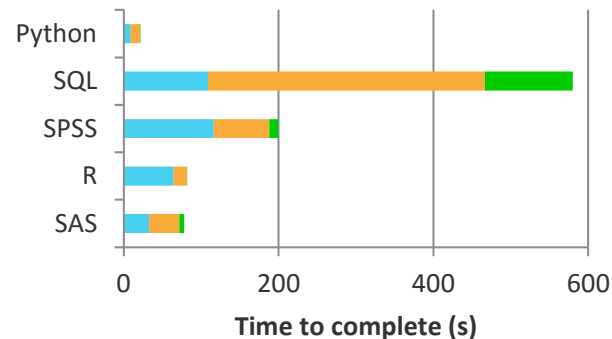


Comparing software

- One of the biggest consideration is how data is stored. SAS scales better
- Software speed tend to reflect its origins
- Feature completeness has improved

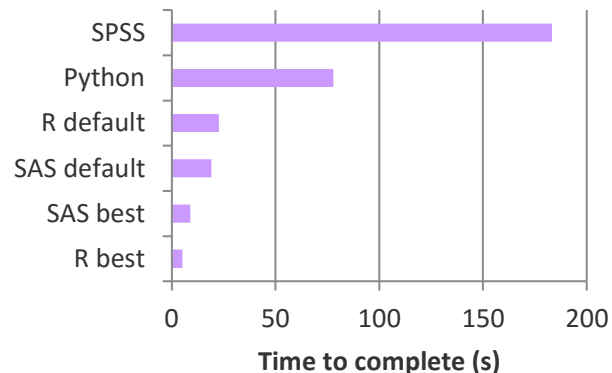
	Hierarchical Clustering	PCA	Decision trees	Boosting with trees	GAM or MARS regression	Mixed models / penalised regression	Neural Networks	SVM
SAS	✓	✓	✓*	*	✓	✓	✓*	✓*
SPSS	✓	✓	✓		✓	✓	✓	✓
R	✓	✓	✓✓	✓✓	✓✓	✓	✓	✓
Python	✓	✓	✓✓	✓✓	*	✓	✓✓	✓

Data handling, 5m rows



■ Merges ■ Sorts ■ Summarise

GLM, 1m rows



Interesting developments in software



Microsoft

REVOLUTION
ANALYTICS

Elasticsearch
HBase
MongoDB
NoSQL
Couchbase
BerkleyDB
Neo4J
Riak
Cassandra

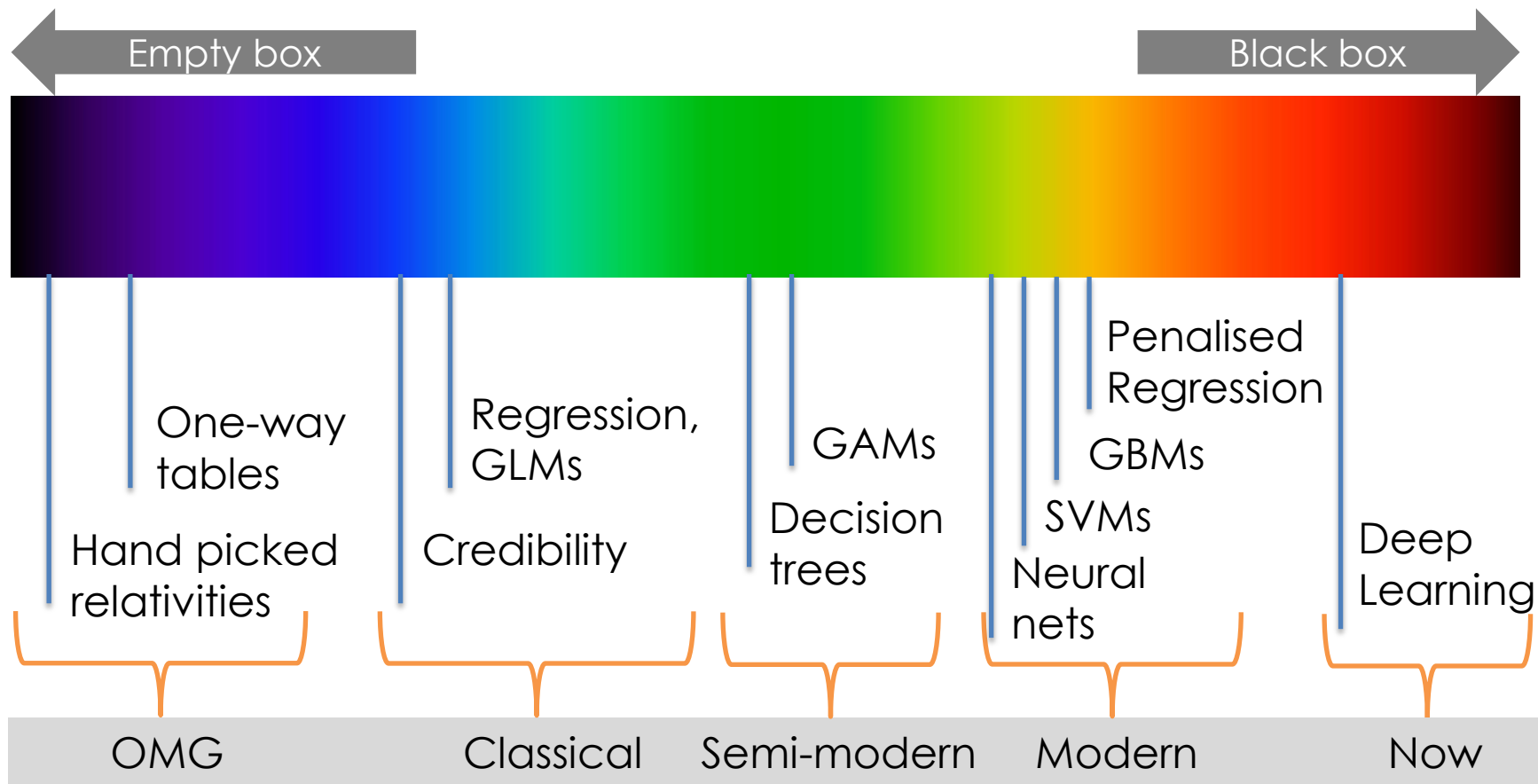


Anaconda



Algorithm choice

(predictive learning)



Interesting developments in algorithms

- Still a lot of mileage in regression framework
- Deep learning goes open source
- We are getting better at 'best practice' (experimental design, model validation etc)
- Hard work remains on the implementation side

Communication choice

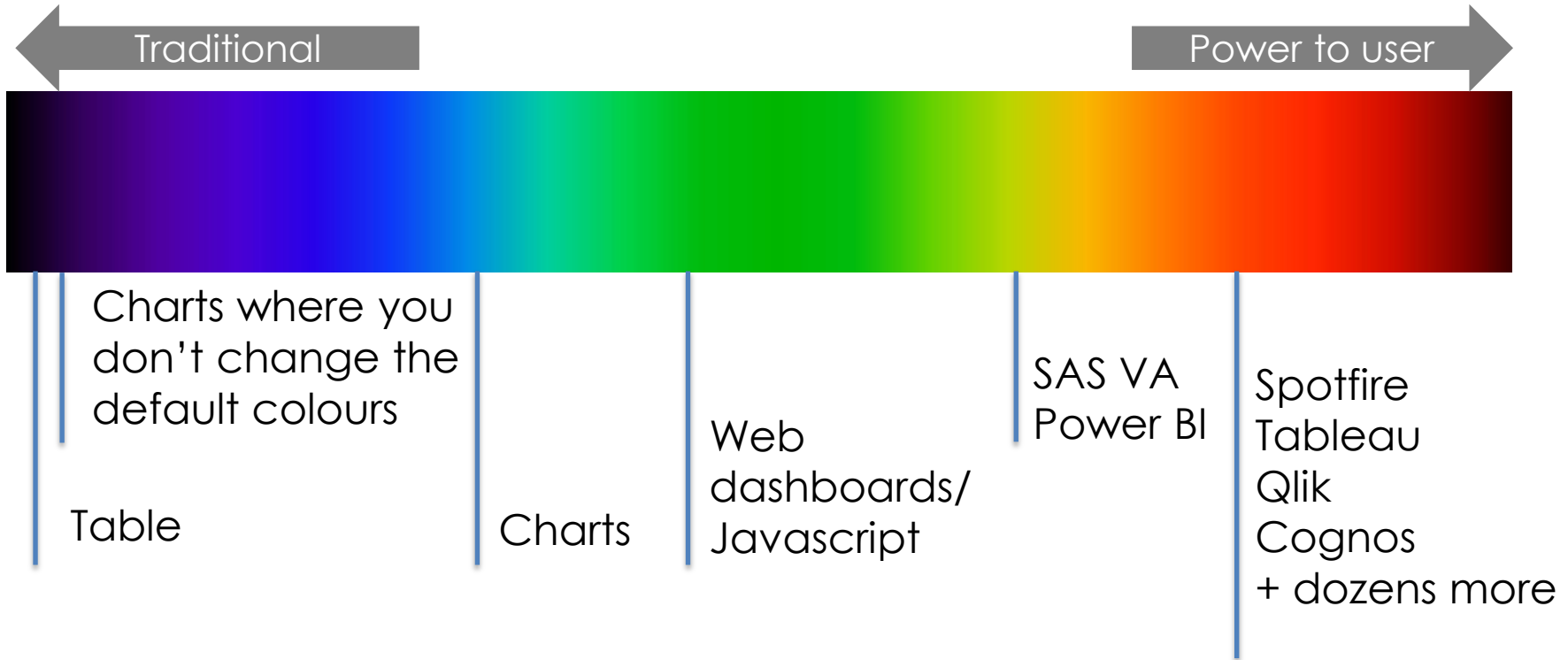
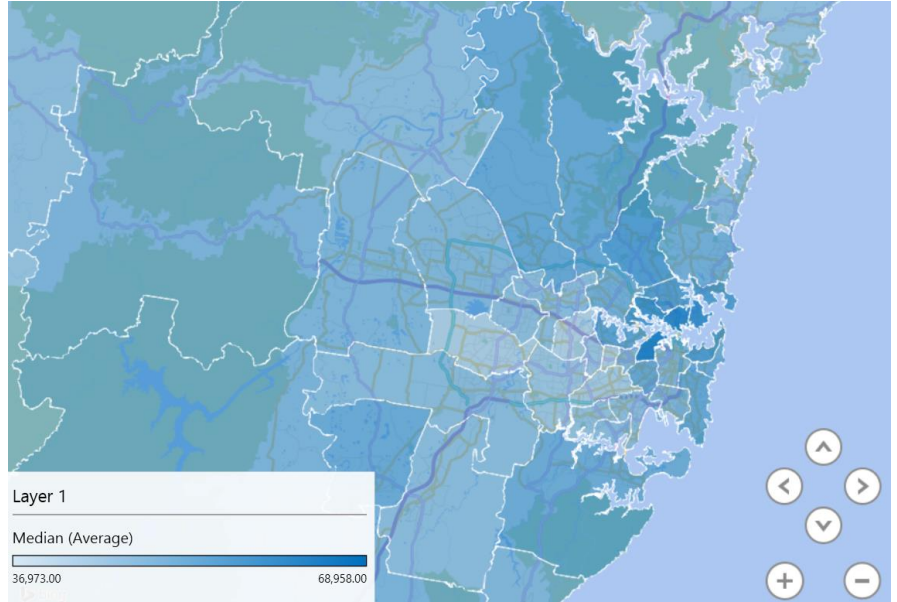


Tableau example:

<https://public.tableau.com/s/gallery/airbnb-prices-san-francisco>

Some useful innovations in communication

- Mapping
- Effect Attribution
- Understanding model uncertainty



Where to next?

- Hardware agnosticism
- Software tend towards lower level, continued standardisation of data
- Algorithms increasingly automatic
- Interactive communication

