

DataScience for Development and Social Change, 2015

Big Data

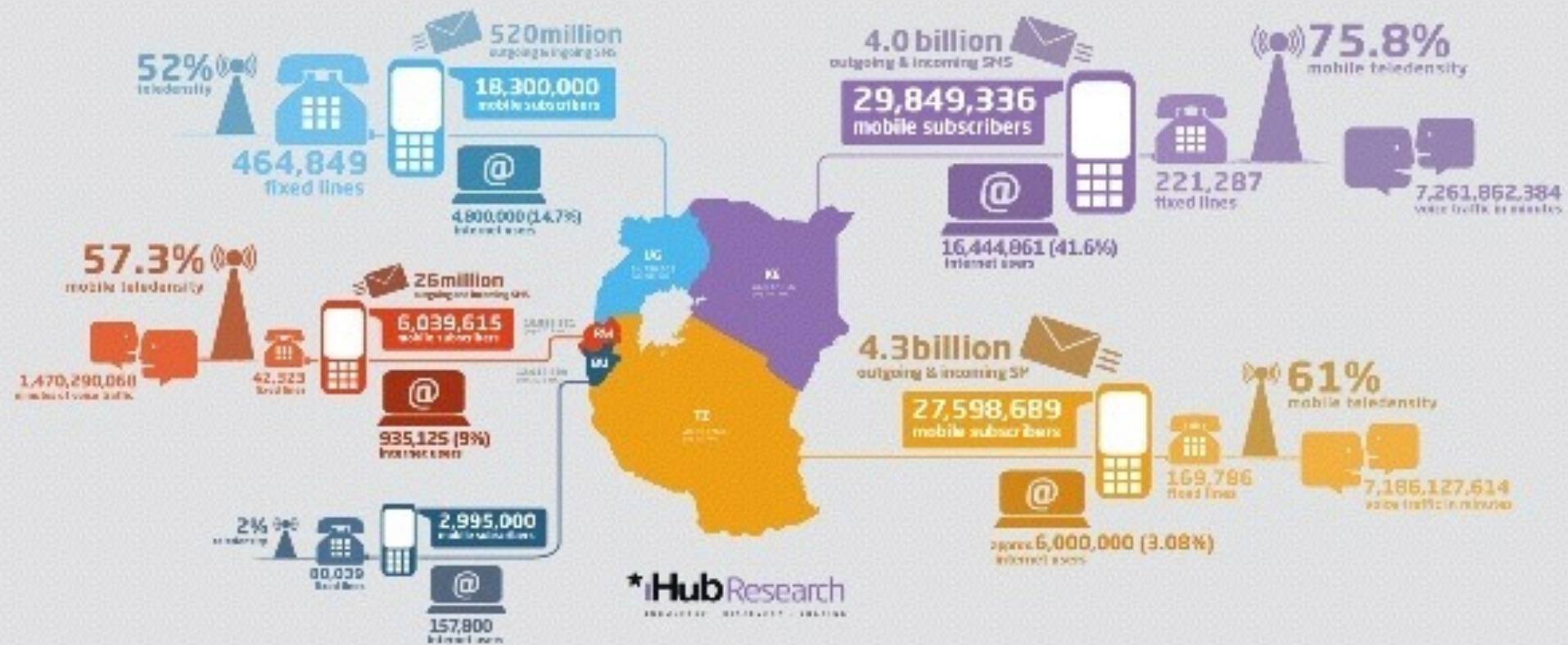
What's special about it, how
to handle it...

“The Firehose”



Volume

Mobile Stats in East Africa



2004-05

[illegible]

77% of patients have the signs of a stroke with age 65 and over.

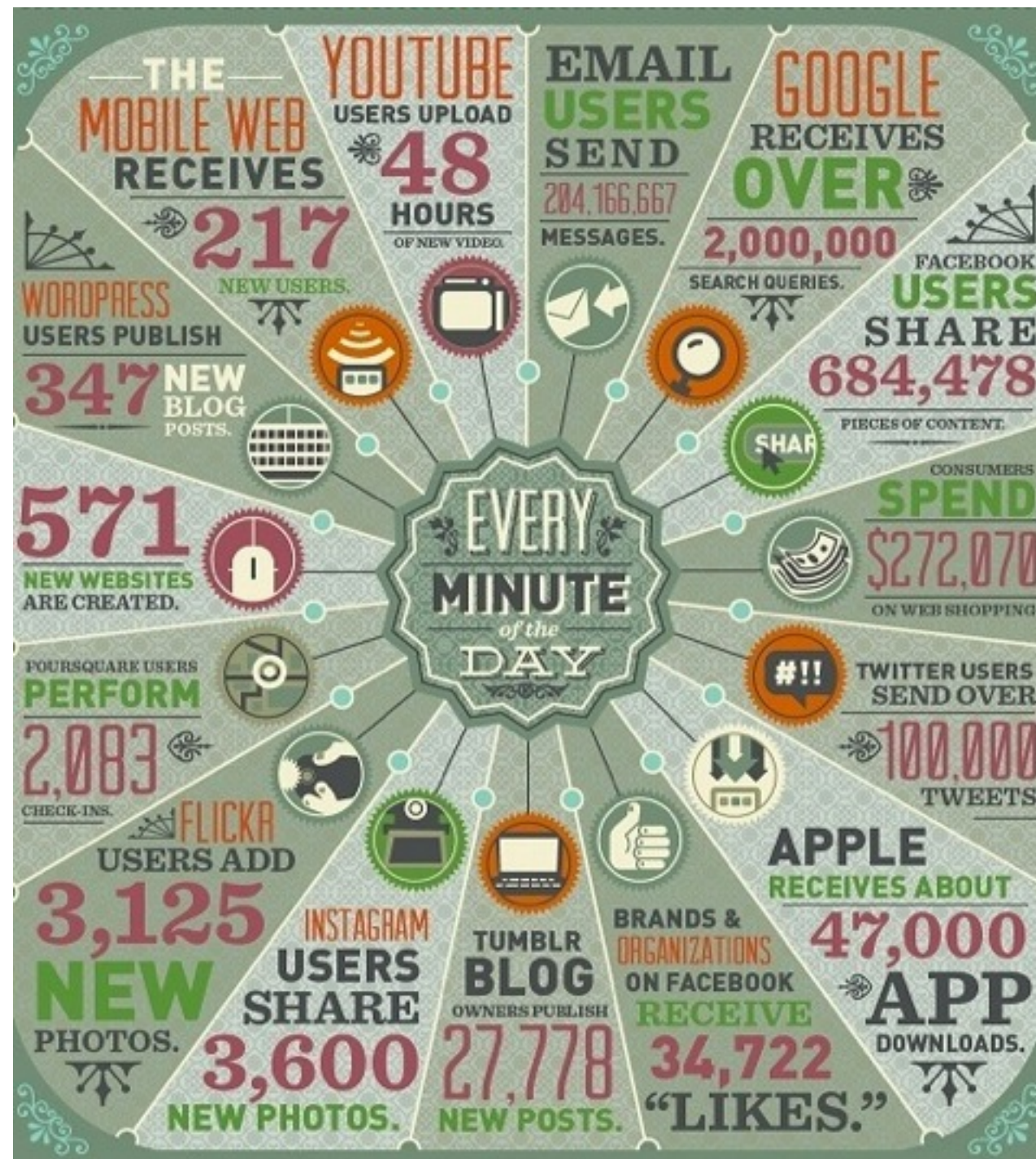
10/2/2005 5:25:55 PM 10/2/2005 5:25:55 PM

[illegible]

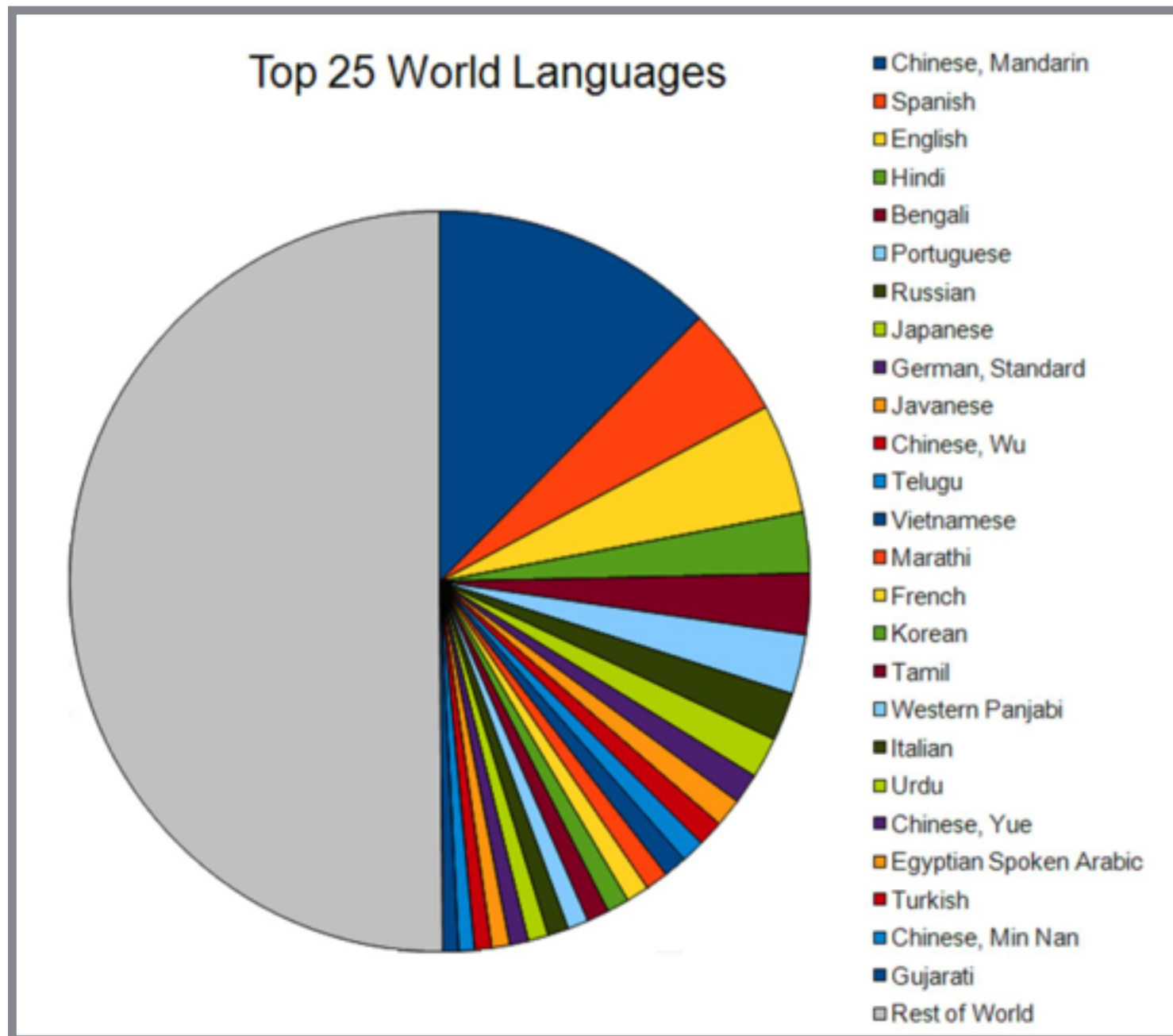
Received 11 October 2014; accepted 11 November 2014

[†] <https://doi.org/10.1093/ajph/101.12.2183>

Velocity



Variety



CSV, json, xml,
excel, pdf,
text,
webpages, rss,
scanned pages,
images, videos,
audiofiles,
maps,
proprietary
formats etc.

Veracity



The “3 other Vs” are Viability, Validation, Verification.

Validation, checking that inputs are real, is a big deal for development data.

Map/Reduce

- ❖ Get Hadoop for VirtualBox: http://www.cloudera.com/content/cloudera/en/downloads/quickstart_vms/cdh-5-3-x.html

Let's go get some data...

- ❖ Twitter API and Twitter Python libraries

And process it

Continuing your Big Data journey
