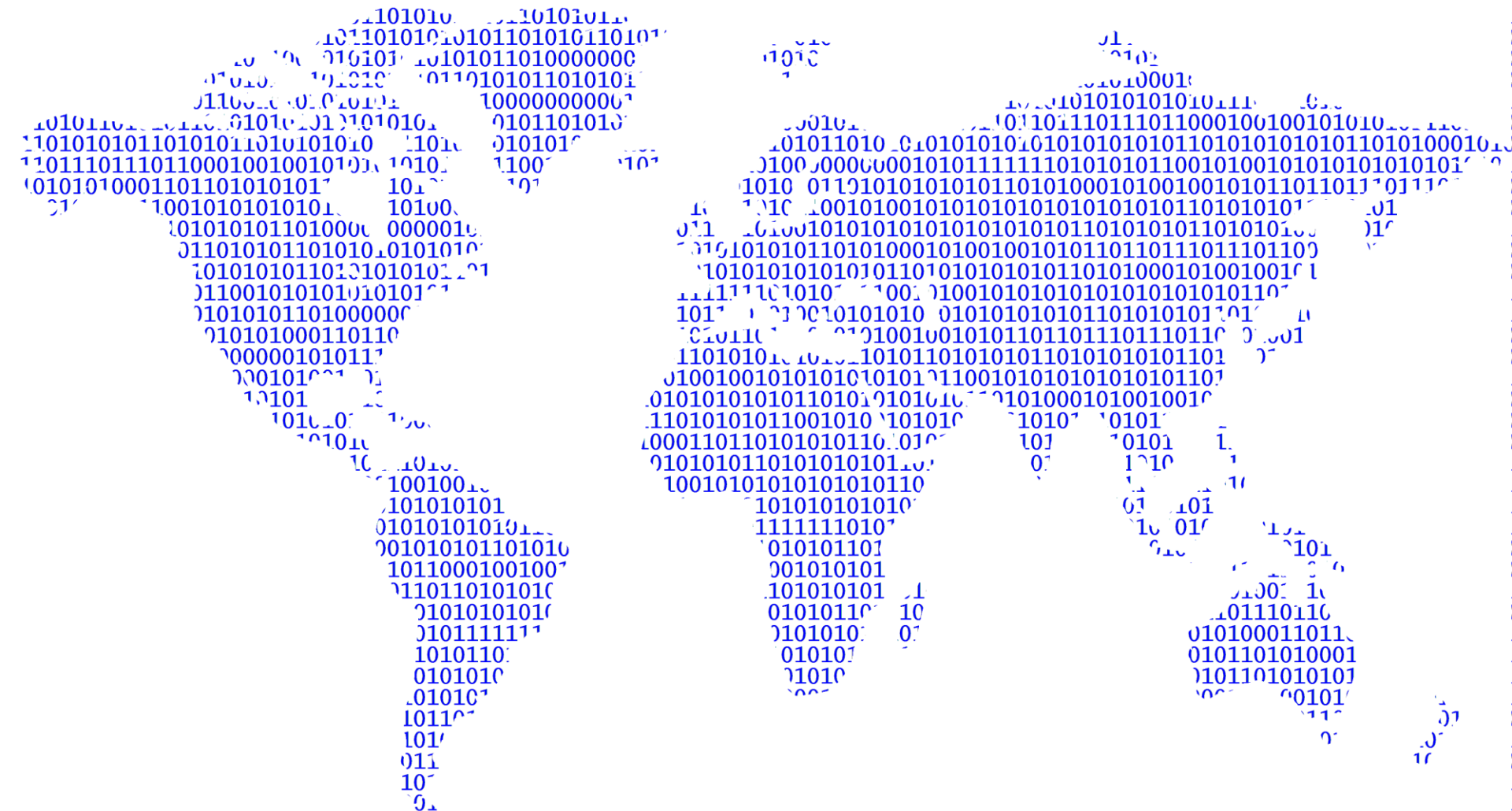
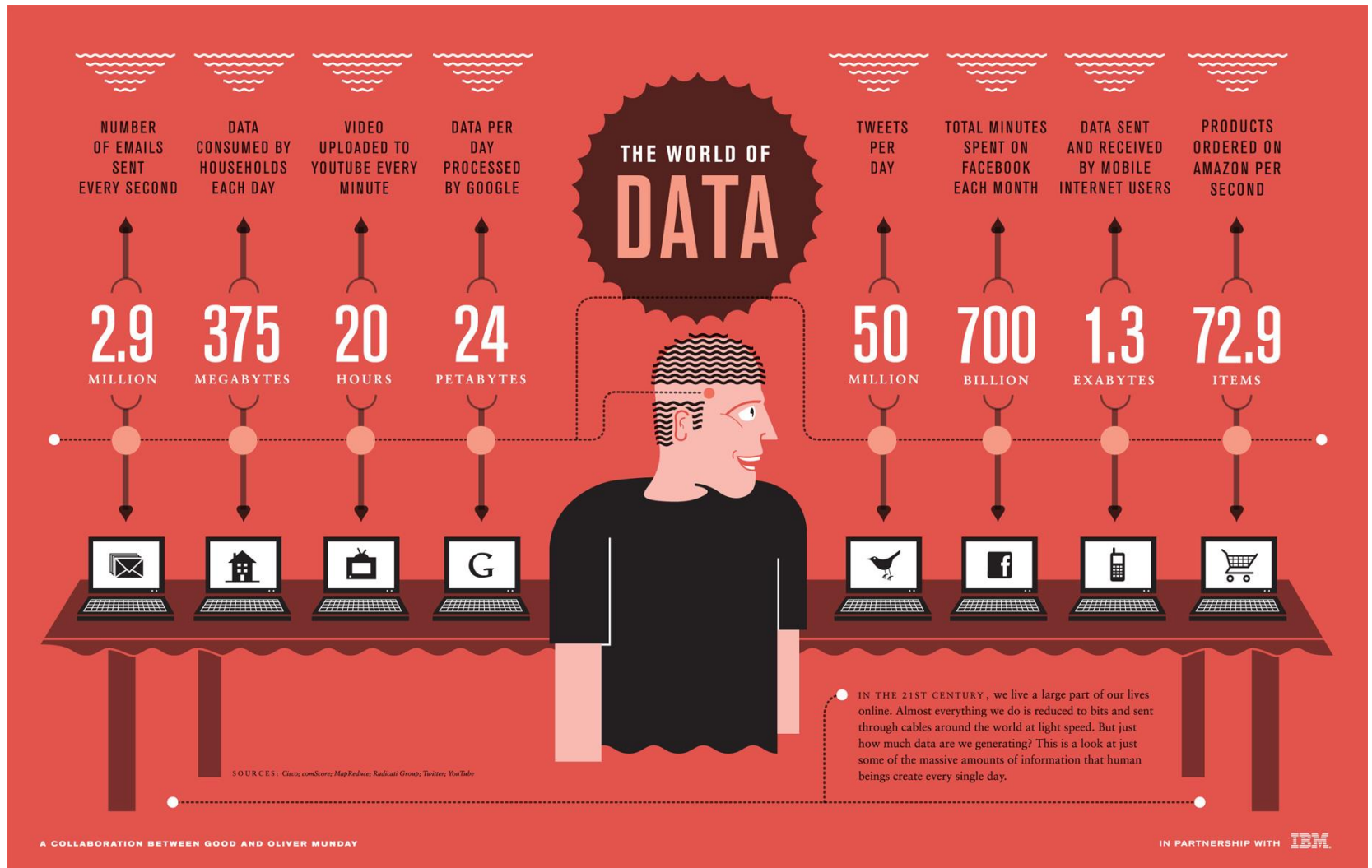


A *Data Based* World



Data growth expectations (by 2010)



Data growth expectations (by 2020)

40 ZETTABYTES

[43 TRILLION GIGABYTES]

of data will be created by 2020, an increase of 300 times from 2005

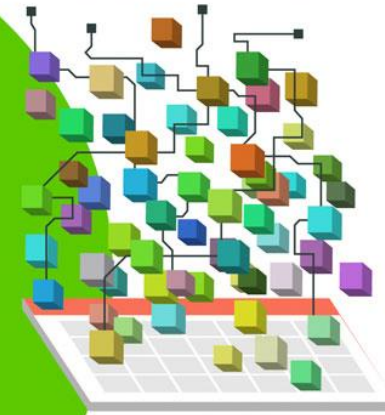


It's estimated that

2.5 QUINTILLION BYTES

[2.3 TRILLION GIGABYTES]

of data are created each day



**Volume
SCALE OF DATA**

**6 BILLION
PEOPLE**
have cell
phones



WORLD POPULATION: 7 BILLION

Most companies in the
U.S. have at least

100 TERABYTES

[100,000 GIGABYTES]

of data stored

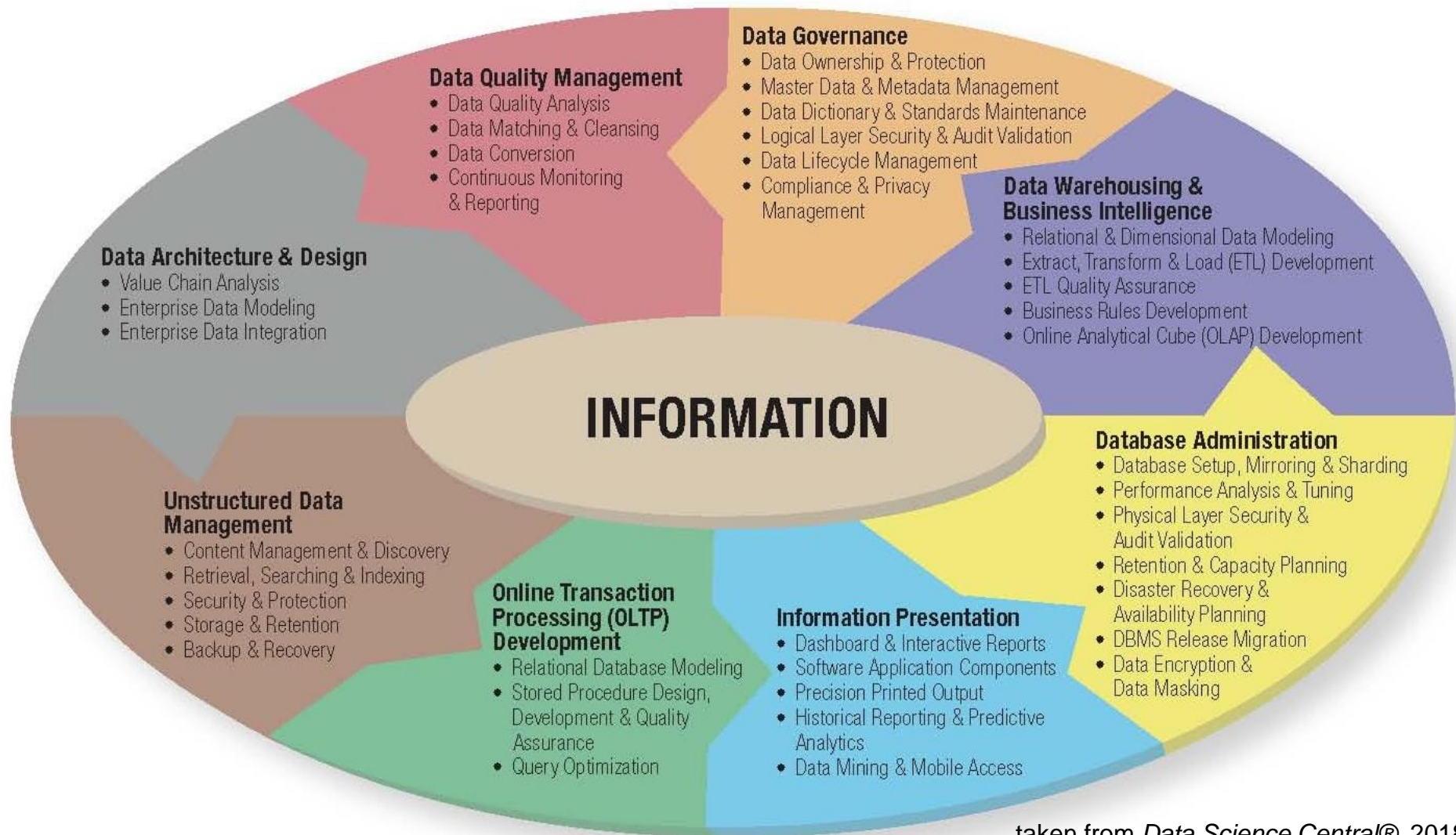
taken from <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>

Why is data so desirable?

Data => Information => Knowledge

Knowledge => Power to:
Choose destiny
Guide people
Free the world

Information: the crucial stage



taken from *Data Science Central*®, 2018

Data related *challenges*

store but also process
all the available data

democratize but also secure
the access to data

merge heterogeneous data but validate
its origin

...

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization: customer tracking and on-site analytics: predictive analytics and econometrics: data warehousing and big data systems: marketing channel insights in Paid Search, SEO, Social, CRM and brand.

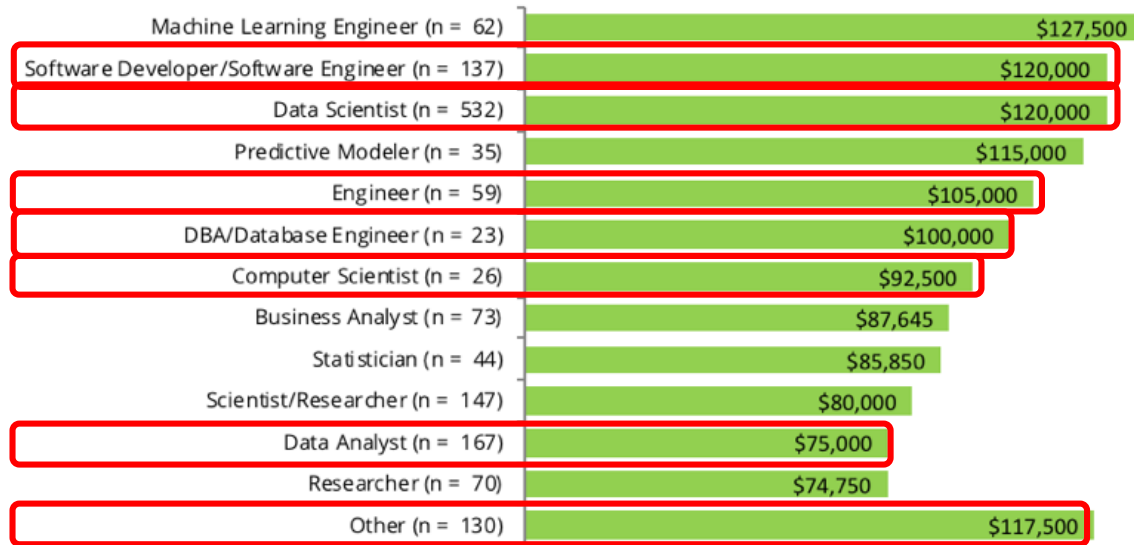
Marketing
DISTILLERY
(c) Krzysztof Zawadzki

Data related challenges

Areas requiring strong data related skills

Data related jobs

Median Annual Compensation by Job Title for United States Data Professionals



Respondents are from the United States. Values are in US Dollars.

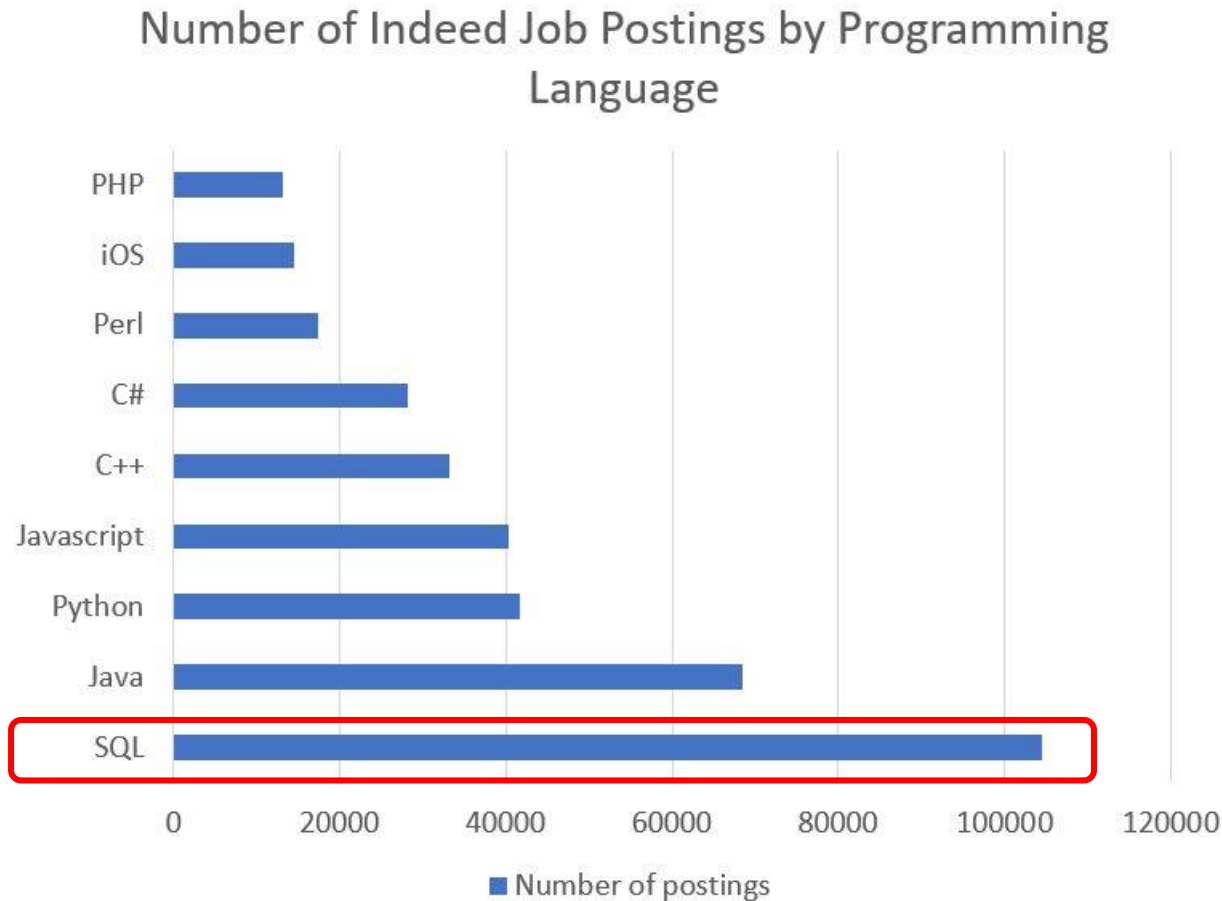
Data are from the Kaggle 2017 The State of Data Science and Machine Learning study. You can learn more about the study and download the data here: <https://www.kaggle.com/surveys/2017>.

Job Titles are ranked by median annual salary. Only job titles with ample sample size (n > 20) are presented.



Copyright 2018 Business Over Broadway

Data and Programming Languages



taken from *Coding Dojo Blog*®, 2017