

# Data Scientist: A Career for 2015 and Beyond

---

Richard F. Eng

PRINCE2, PMP, CSQE, CRE, CQE

[reng@mitre.org](mailto:reng@mitre.org)

703-201-9112

20 March 2015

*The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author.*

# Outline

---

- **Hype or Reality?**
- **Data Scientist**
- **Opportunities**
- **Unicorns**
- **Acquiring Unicorn Skills**
- **Challenges**
- **Conclusions**
- **Next Steps**

# Hype or Reality?

- **“Data Scientist: The Sexiest Job of the 21st Century”**
  - Thomas H. Davenport and D.J. Patil
- **“Analytics is defined as the scientific process of transforming data into insight for making better decisions.”**
  - The Institute for Operations Research and the Management Sciences (INFORMS)
- **“With more and more companies using big data, the demand for data analytic specialists,—sometimes called data scientists, who know how to manage the tsunami of information, spot patterns within it and draw conclusions and insights—is nearing a frenzy.”**
  - Chris Morris, CNBC

# Data Scientist

- **“A person who is better at statistics than any software engineer and better at software engineering than any statistician.”**

**<https://www.youtube.com/watch?v=O6kZkq3rdwc>**

**Josh Wills**

**Director of Data Science at Cloudera**

- **“Data scientists are inquisitive: exploring, asking questions, doing “what if” analysis, questioning existing assumptions and processes. Armed with data and analytical results, a top-tier data scientist will then communicate informed conclusions and recommendations across an organization’s leadership structure. ”**

**<http://www-01.ibm.com/software/data/infosphere/data-scientist/>**

**Anjul Bhambhri, IBM**

# Opportunities

## Job Trends from Indeed.com

— Big Data Analytics



Source: <http://cdn.edureka.co/blog/wp-content/uploads/2015/01/trends-analytics1>

# What Do Data Scientist Do?

- **Interface with analytics, product management, and operations teams.**
- **Perform large-scale data analysis and develop effective statistical models** for segmentation, classification, optimization, time series, etc.
- **Design and implement reporting dashboards** that track **key business metrics** and provide **actionable insights**
- **Identify actionable insights, suggest recommendations and influence the direction** of the business by effectively communicating results to cross functional groups
- **Work closely with Product or Engineering & Operations teams** to **proactively** create rule and manage decisions
- **Suggest improvements** in the tools and techniques to help scale the team

# What Do Data Scientists Do? (cont.)

---

- Apply data-mining, machine learning and/or graph analysis techniques for a variety of modeling and relevance problems involving users, their relationships, their tweets and their interests.
- Design and evaluate novel approaches for handling high-volume real-time data streams.
- Code using primarily Java, Scala, and scripting languages such as Python or Ruby.
- Conduct design and code reviews.
- Work with large unstructured and structured data sets (*multi-terabyte+, 100MM+ daily transaction volumes*).
- Utilize data science and quantitative methodologies to help shape clinical care and long-term planning

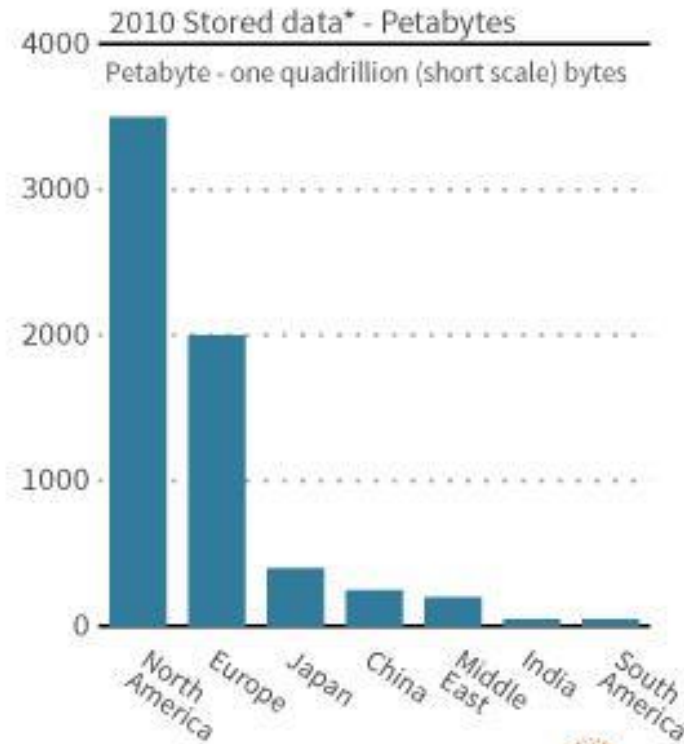
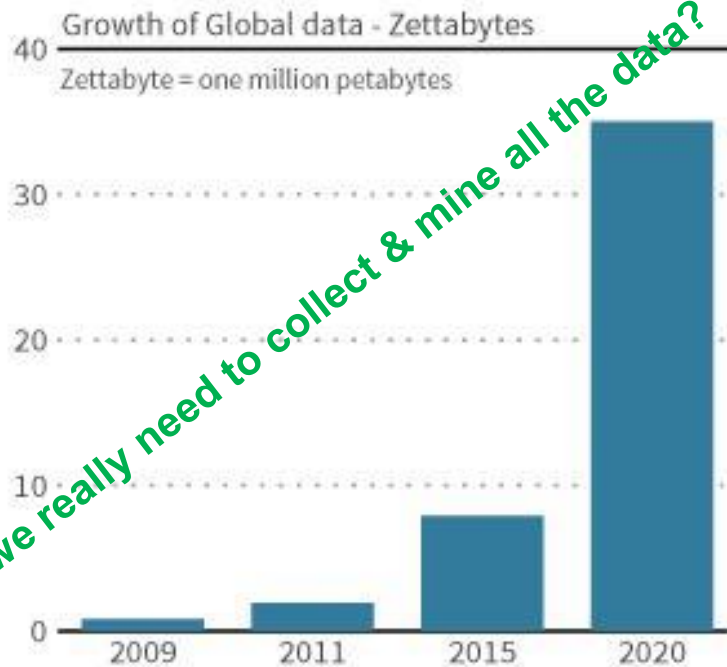
# Finding Real Insights ...

## Data Sources

- Sensor data
- Log data
- Internet of Things
- Etc.

## Big data growth

Big data market is estimated to grow 45% annually to reach \$25 billion by 2015



\*greater than

Sources: Nasscom -CRISIL GR&A analysis

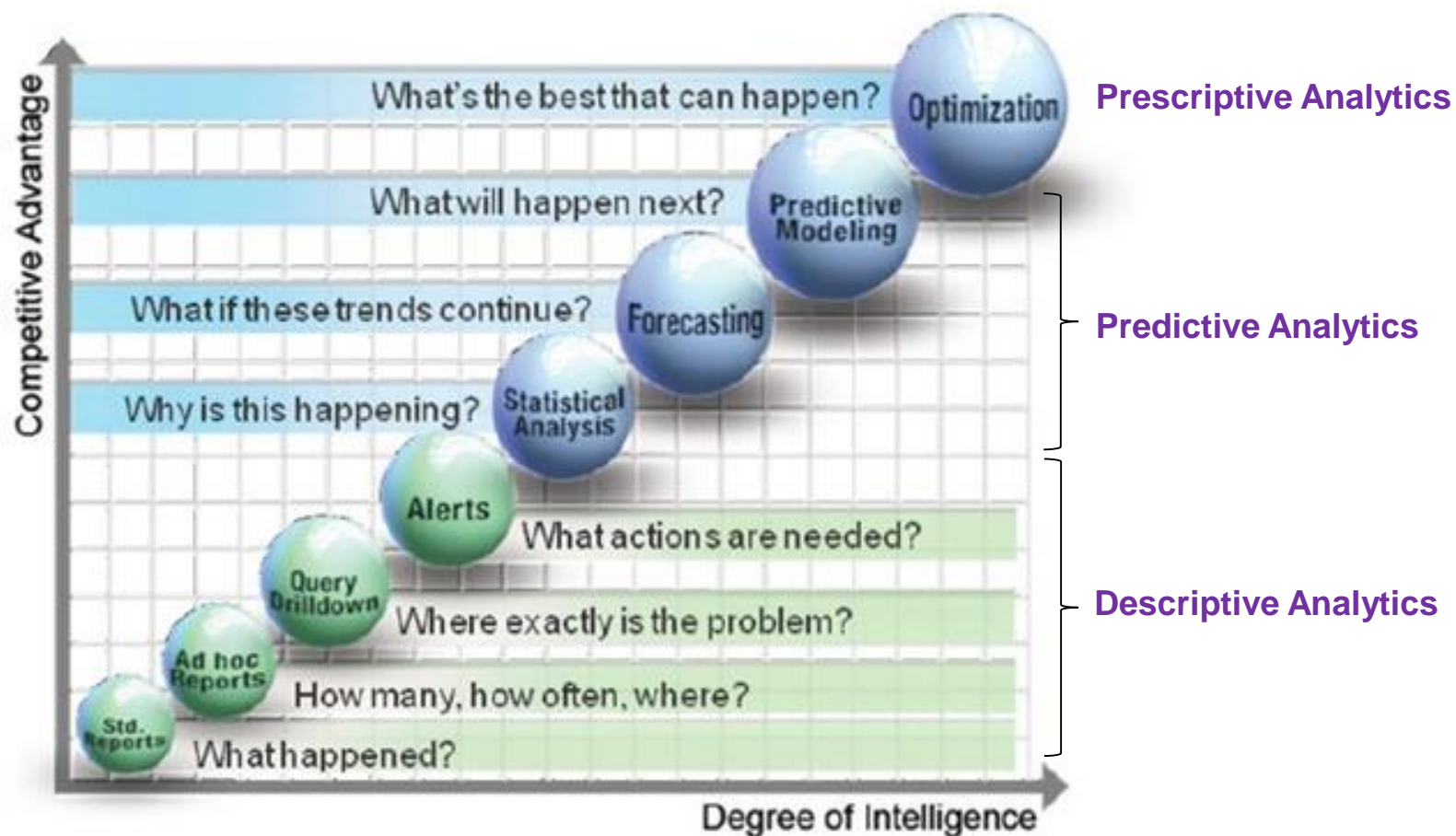


Reuters graphic/Catherine Trevethan 05/10/12

Do we really need to collect & mine all the data?

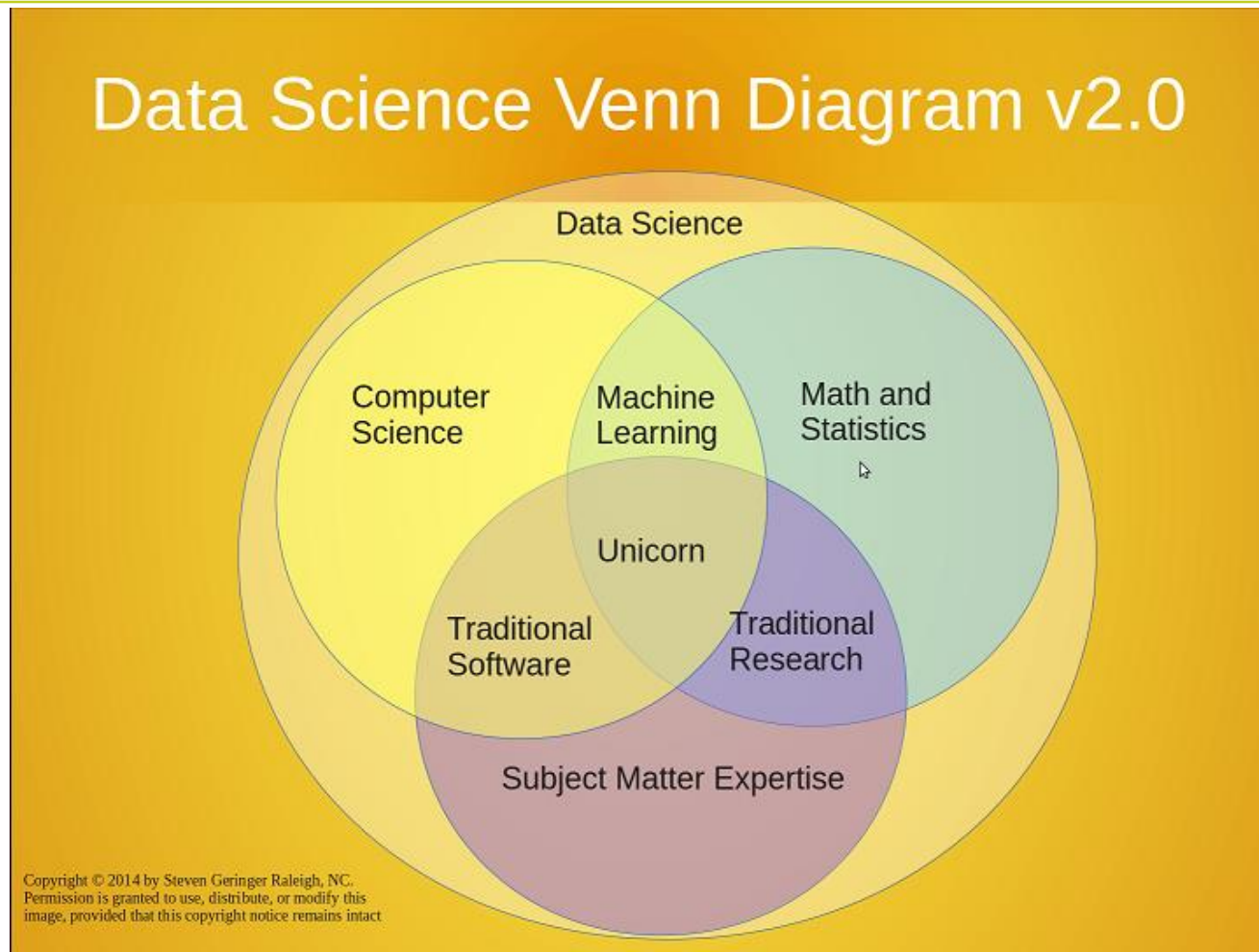


# Data Analytics



*Optimization helps you determine the best that can happen, so you can take action in ways that will deliver significant performance improvements. Advances in technology have made this process easier and more powerful.*

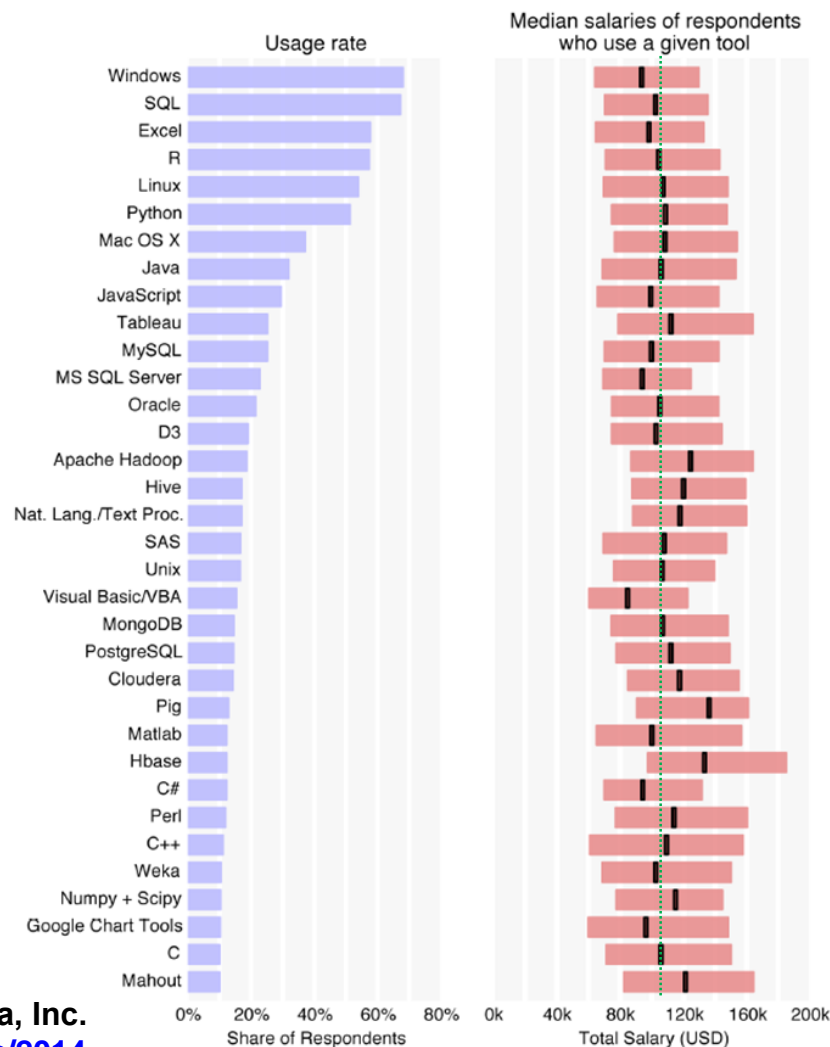
# Unicorns



Source: [http://insidebigdata.com/wp-content/uploads/2014/01/Venn.png?utm\\_content=buffer21f7d&utm\\_medium=social&utm\\_source=twitter.com&utm\\_campaign=buffer](http://insidebigdata.com/wp-content/uploads/2014/01/Venn.png?utm_content=buffer21f7d&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer)

# Unicorn Skills

- Communications
- Leadership
- Domain
- Business
- Tools
  - R
  - Python
  - Java
  - Tableau
  - SQL
  - Hadoop
  - SAS/SAS Enterprise Miner



Source: 2014 Data Science Salary Survey. O'Reilly Media, Inc. CA. Retrieved from <http://www.oreilly.com/data/free/files/2014-data-science-salary-survey.pdf>

# Acquiring Unicorn Skills

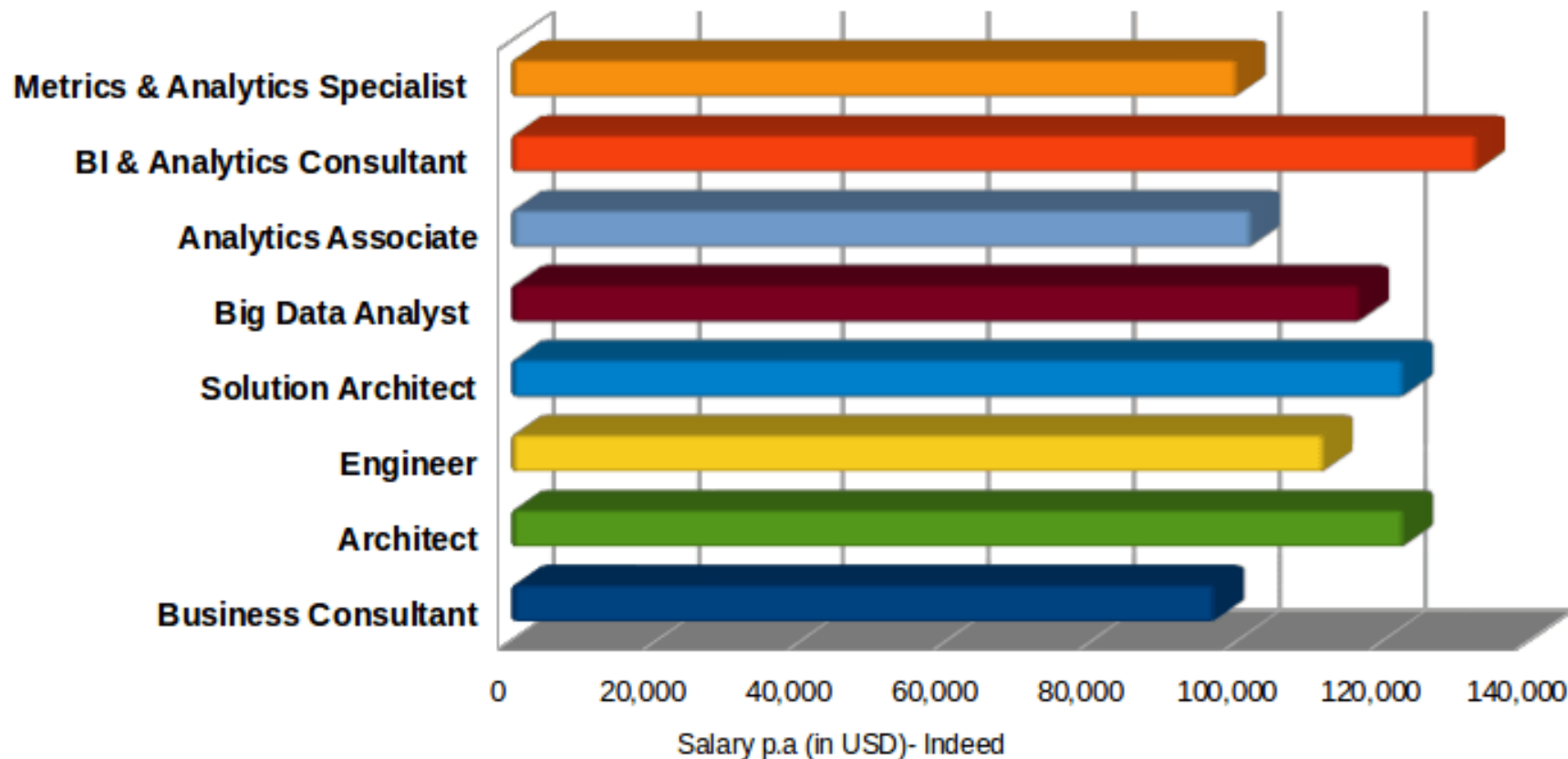
---

- **Self-study**
- **On the job training**
- **Websites**
  - Data Science Central <http://www.datasciencecentral.com/>
  - Kaggle <http://www.kaggle.com/>
  - KDnuggets <http://www.kdnuggets.com/>
- **Journals**
  - IEEE Transactions on Computational Intelligence and AI in Games
  - IEEE Computational Intelligence
- **Conferences**
  - IEEE International Conference on Big Data 2014
  - IEEE 2014 International Conference on Data Science & Engineering (ICDSE)
  - Predictive Analytics World
- **Courses and Degree programs**  
<http://www.kdnuggets.com/education/online.html>

# Salaries

Note: Salary data varies among surveys

## Big Data Analytics Job Titles & Salaries



Source: <http://cdn.edureka.co/blog/wp-content/uploads/2015/01/salaryjobtitles.png>

# Challenges

---

- **Lots of data NOT ALWAYS EQUAL TO Insights**
  - Need to carefully select attributes
  - Need to understand the context of the problem and results
  - Do the model results make sense!
  - Realize that models degrade with time!
- **Gaining**
  - Data analytics experience
  - Knowledge and skills
  - Experience that employers care about
- **Employers**
  - Finding skilled staff
  - Staff in Centralized Data Analytics Center versus Business Units
  - Retaining staff
  - Training and growing staff

# Conclusions

---

- **Electrical Engineers, Computer Science, and Math Professionals can transition to Data Science**
- **Already know and use various aspects of**
  - Descriptive analytics
  - Prescriptive analytics
  - Predictive analytics
- **Resources exist to allow information technology professionals to fill-in their knowledge gaps**
- **Data Science is a growth area**

## Next Steps

---

- Take courses and read to fill in the knowledge gaps
- Add a data science, data analytics, or predictive analytics degree if time permits
- Find small projects where you are to
  - Leverage your current skills and strengths
  - Apply data analytics as a value added benefit
- **Talk with people interested in data science and data analytics**

*“Continuous effort – not strength or intelligence – is the key to unlocking our potential.”*

*Winston Churchill*



# Back-up Slides

---

# Resources and Further Reading

- Davenport, T. H., and Kim, J. (2013). Keeping up with the quants: Your guide to understanding and using analytics. Boston, MA: Harvard Business Review Press.
- Edureka! (2015). 10 reasons why big data analytics is the best career move. Retrieved from <http://www.edureka.co/blog/10-reasons-why-big-data-analytics-is-the-best-career-move>
- Few, S. (2009). Now you see it: Simple visualization techniques for quantitative analysis. Oakland, CA: Analytic Press.
- Fisher, A. (2013). Big Data could generate millions of new jobs. Fortune. Retrieved from <http://fortune.com/2013/05/21/big-data-could-generate-millions-of-new-jobs/>
- Kdnuggets – Data Mining Community  
<http://www.kdnuggets.com/2014/02/best-tweets-in-january.html>
- Siegel, E. (2013). The power to predict who will click, lie, buy, or die. Wiley.
- The Data Warehousing Institute (TDWI) <http://tdwi.org/Home.aspx>

# Resources and Further Reading (cont.)

---

## ■ Journals

- IEEE Transactions on Computational Intelligence and AI in Games  
<http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=4804728>
- IEEE Computational Intelligence  
<http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=33585>

## ■ Conferences

- IEEE International Conference on Big Data 2014  
<http://cci.drexel.edu/bigdata/bigdata2014/>
- IEEE 2014 International Conference on Data Science & Engineering (ICDSE)  
[http://www.ieee.org/conferences\\_events/conferences/conferencedetails/index.html?Conf\\_ID=32781](http://www.ieee.org/conferences_events/conferences/conferencedetails/index.html?Conf_ID=32781)
- Predictive Analytics World <http://www.predictiveanalyticsworld.com/>

# Courses and Certification Programs (Sample)

---

- CalTech, Learning from Data, <http://work.caltech.edu/telecourse.html>
- Coursera <https://www.coursera.org/specialization/jhudatascience/1/overview>
- MIT Open Courseware <http://ocw.mit.edu/courses/sloan-school-of-management/15-075j-statistical-thinking-and-data-analysis-fall-2011/>
- New Jersey Institute of Technology Certificate in Data Mining <http://online.njit.edu/programs/certs/datamining-cert.php>
- Stanford Data Mining and Applications Graduate Certificate <http://scpd.stanford.edu/public/category/courseCategoryCertificateProfile.do?method=load&certificateId=1209602>
- Statistics.com <http://www.statistics.com/landing-page/data-analytics-courses/>

# Degrees Programs (Sample)

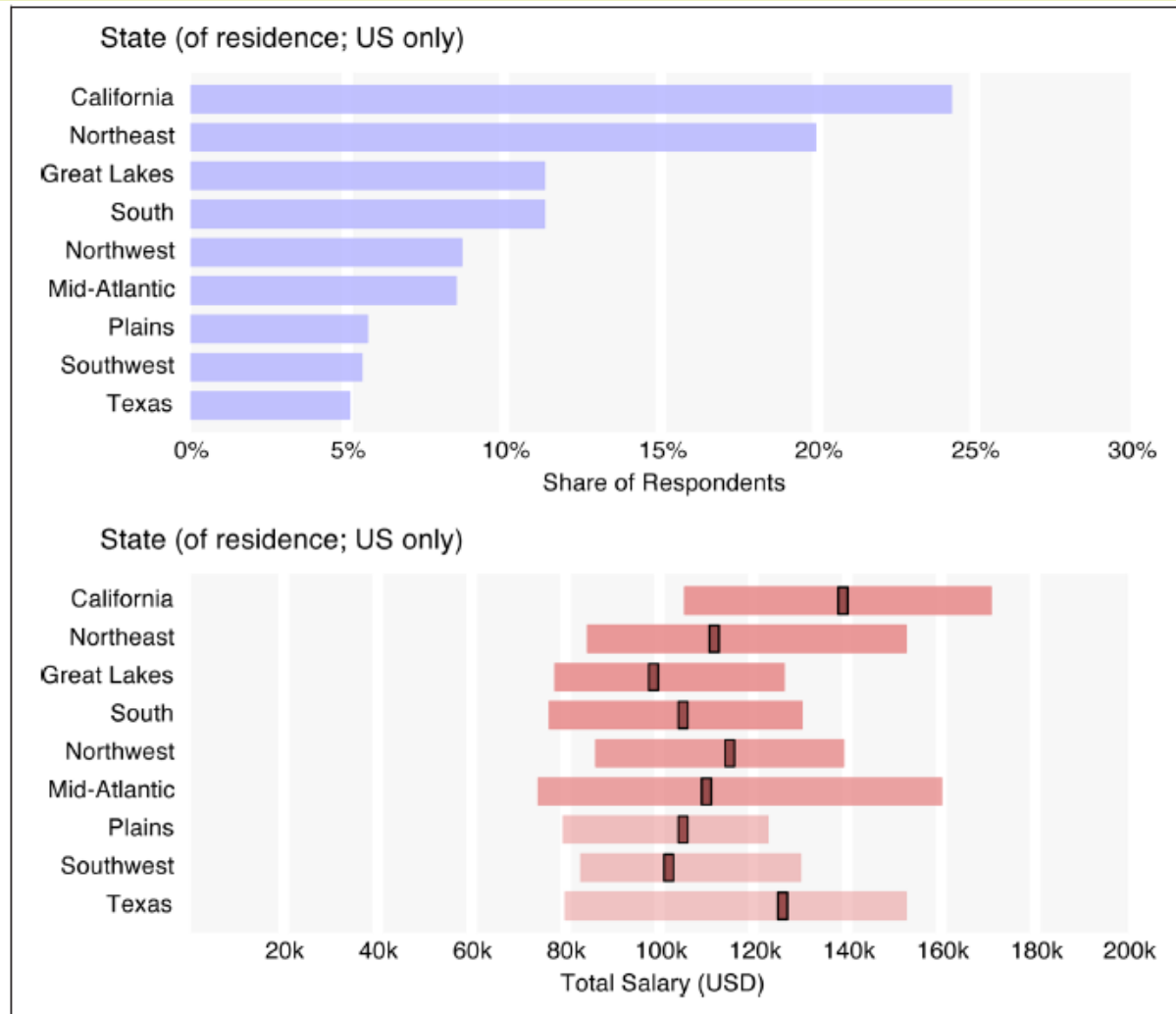
---

- **Berkeley Masters of Information and Data Science**  
<http://datascience.berkeley.edu/>
- **DePaul University, Master of Science Predictive Analytics**  
<http://www.cdm.depaul.edu/academics/Pages/MS-In-Predictive-Analytics.aspx>
- **Northwestern University, Master of Science in Predictive Analytics**  
[http://sps.northwestern.edu/info/predictive-analytics.php?utm\\_source=KDNuggets\\_TextLink&utm\\_medium=TextLink&utm\\_term=FY14&utm\\_content=MSPA&utm\\_campaign=MSPA\\_KDnuggTL14&src=kdnugg\\_text\\_fy14](http://sps.northwestern.edu/info/predictive-analytics.php?utm_source=KDNuggets_TextLink&utm_medium=TextLink&utm_term=FY14&utm_content=MSPA&utm_campaign=MSPA_KDnuggTL14&src=kdnugg_text_fy14)
- **NYU Stern, Master of Science in Business Analytics** [http://web-marketing.stern.nyu.edu/global-programs/business-analytics/?campaign\\_id=701d0000000Sfo5&utm\\_medium=email&utm\\_source=kdnuggets&utm\\_term=dedicatedemail](http://web-marketing.stern.nyu.edu/global-programs/business-analytics/?campaign_id=701d0000000Sfo5&utm_medium=email&utm_source=kdnuggets&utm_term=dedicatedemail)
- **University of Maryland, Masters of Science in Data Analytics**  
<http://www.umuc.edu/analytics/index.cfm>

# Common Job Posting Terms



# Salaries by State



Source: 2014 Data Science Salary Survey. O'Reilly Media, Inc. CA. Retrieved from <http://www.oreilly.com/data/free/files/2014-data-science-salary-survey.pdf>

# Factors Influencing Salaries

Source: 2014 TDWI Salary, Roles, and Responsibilities Report

## AVERAGE SALARY BY YEARS AT CURRENT COMPANY

	2009	2010	2011	2012	2013	Respondents*
0–1 years	\$102,176	\$109,645	\$109,236	\$104,370	\$106,973	21%
2–3 years	\$102,073	\$101,904	\$103,585	\$109,635	\$103,932	26%
4–5 years	\$97,251	\$98,217	\$99,994	\$109,418	\$107,349	12%
6–10 years	\$101,092	\$102,508	\$106,437	\$105,028	\$106,874	22%
11–20 years	\$101,503	\$103,606	\$110,045	\$109,932	\$109,399	16%
20 or more years	\$104,380	\$107,888	\$114,350	\$110,127	\$111,067	4%

People who moved tended to make more money.

Having 5+ years of experience pays off.

## AVERAGE SALARY BY YEARS OF BI/DW EXPERIENCE

	2009	2010	2011	2012	2013	Respondents*
0–1 years	\$82,170	\$86,769	\$83,941	\$86,347	\$79,822	6%
2–3 years	\$90,307	\$89,287	\$89,837	\$84,715	\$93,164	14%
4–6 years	\$92,224	\$92,046	\$99,732	\$98,967	\$94,178	19%
7–9 years	\$101,753	\$102,820	\$103,454	\$104,193	\$103,953	17%
10 or more years	\$113,784	\$117,205	\$118,512	\$120,376	\$120,580	44%

\*2013 data

## AVERAGE SALARY BY CERTIFICATION

	2009	2010	2011	2012	2013	Respondents*
0 certifications	\$99,396	\$100,816	\$105,985	\$105,253	\$103,439	58%
1 certification	\$98,604	\$102,895	\$100,880	\$107,688	\$104,984	16%
2 certifications	\$103,547	\$105,838	\$109,113	\$108,133	\$112,219	14%
3 certifications	\$108,445	\$102,543	\$115,466	\$115,203	\$111,256	6%
4+ certifications	\$109,293	\$115,381	\$104,466	\$117,161	\$116,005	6%

\*2013 data

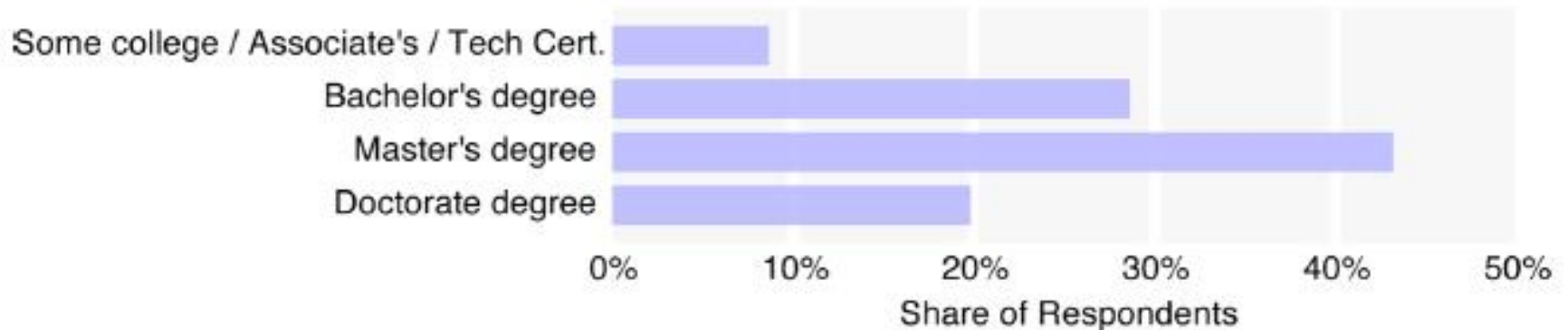
Having IT certifications increases salaries.

Most certifications require at least 5 years of professional experience.

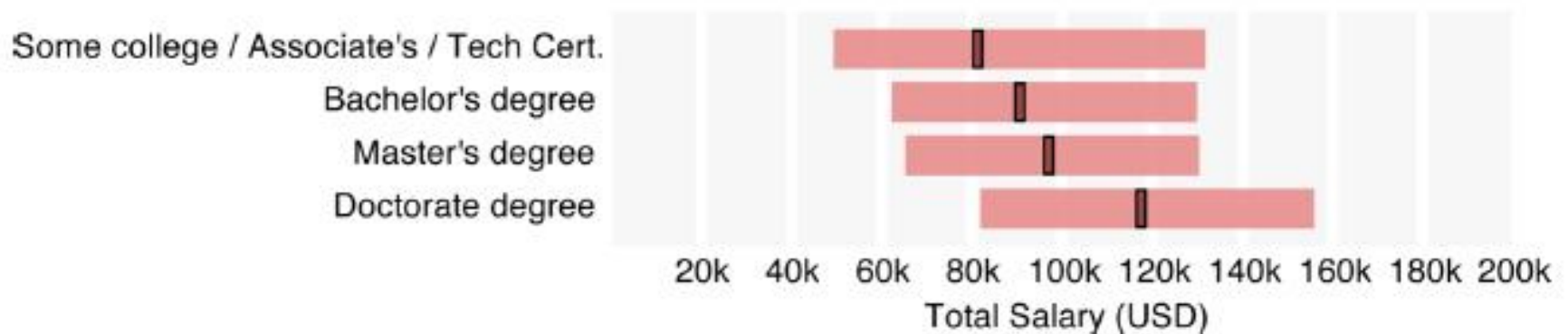


# Impact of Education

## Education (highest level attained)

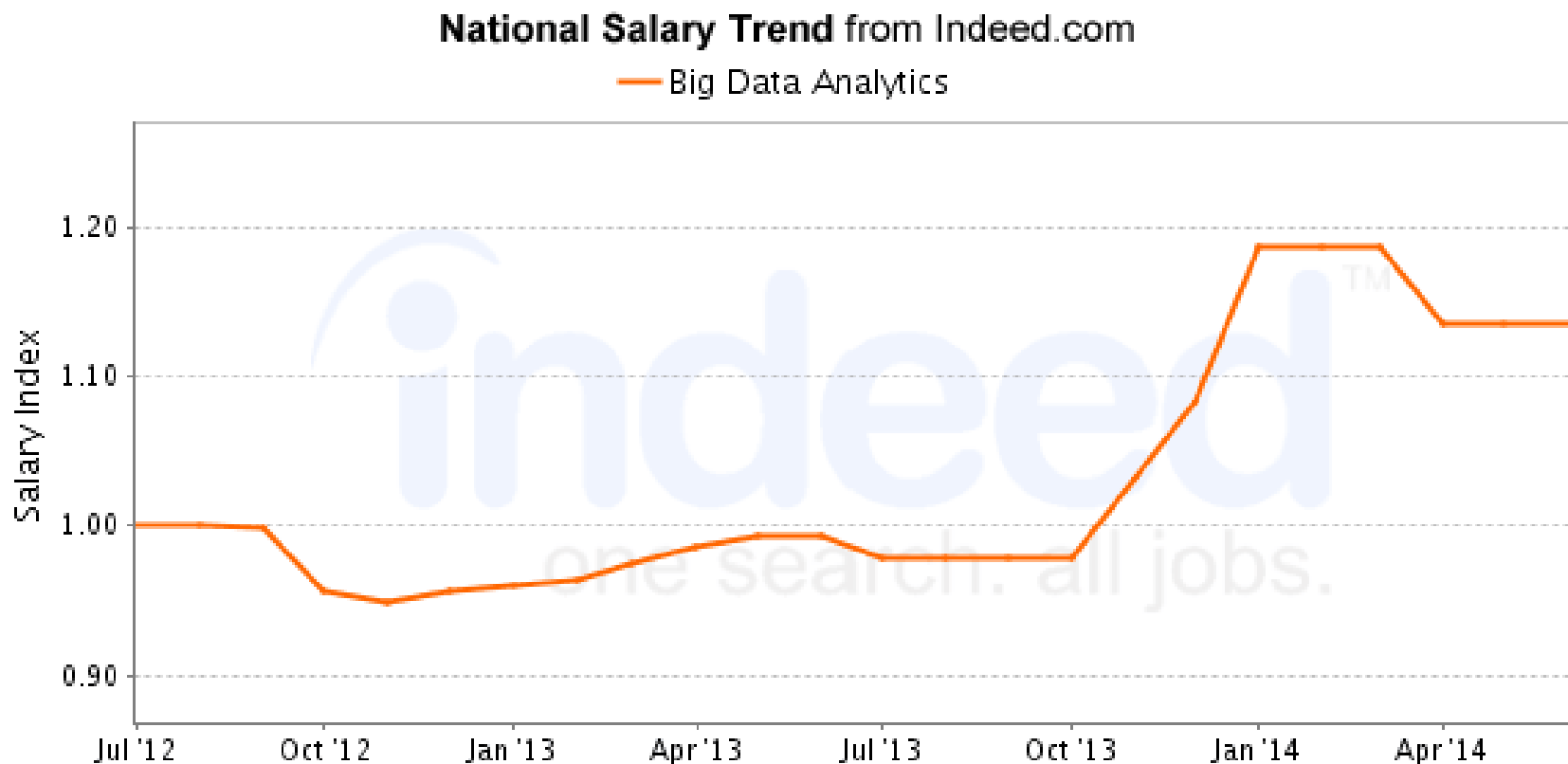


## Education (highest level attained)



Source: 2014 Data Science Salary Survey. O'Reilly Media, Inc. CA. Retrieved from <http://www.oreilly.com/data/free/files/2014-data-science-salary-survey.pdf>

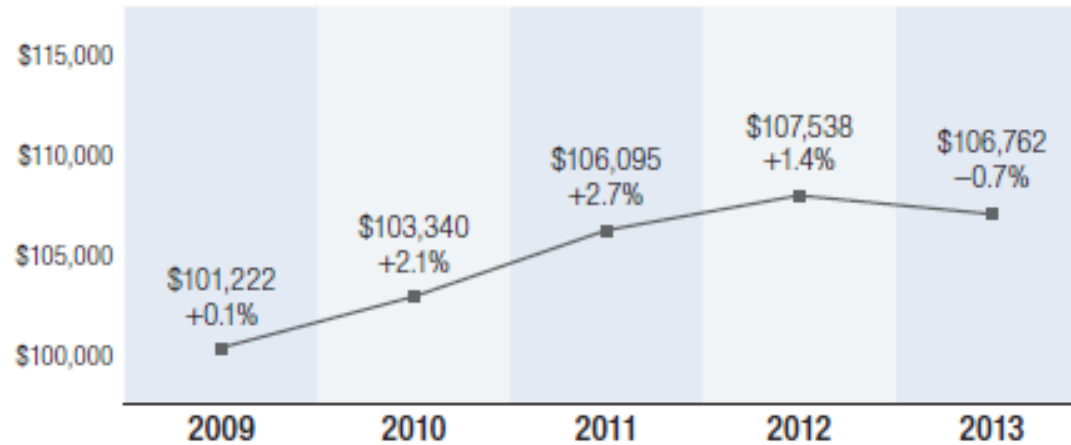
# Salary Trends



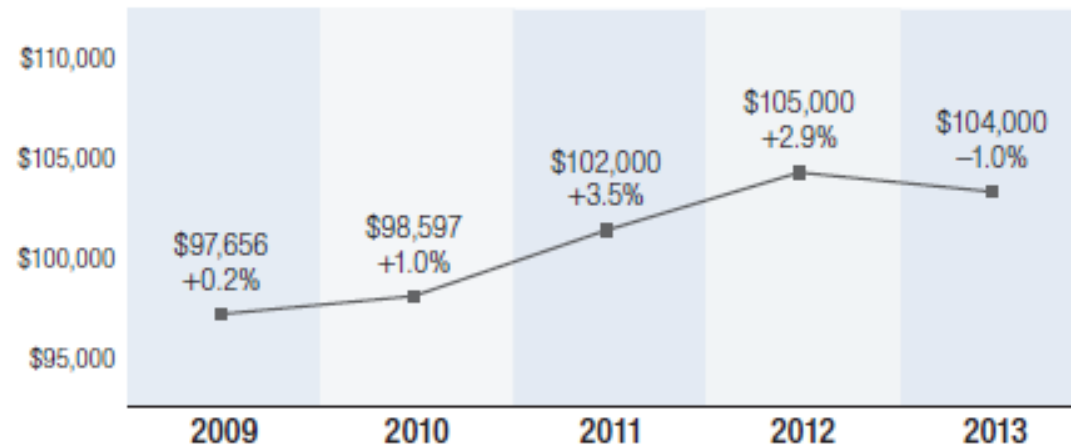
Source: <http://cdn.edureka.co/blog/wp-content/uploads/2015/01/salary-1.png>

# Salaries

## AVERAGE (MEAN) SALARIES



## MEDIAN SALARIES



# Opportunities

## AVERAGE SALARY BY REGION

	2011	2012	2013	Respondents*
Mid-Atlantic	\$116,597	\$135,568	\$122,258	6%
Southwest	\$110,116	\$118,487	\$118,661	6%
Pacific	\$113,552	\$117,649	\$118,396	14%
Northeast	\$111,659	\$110,867	\$114,333	15%
Southeast	\$101,633	\$109,630	\$106,108	11%
South	\$104,038	\$104,692	\$106,063	4%
Midwest	\$100,939	\$99,006	\$102,823	21%
Rocky Mountains	\$105,294	\$106,308	\$100,985	4%
Central Plains	\$106,786	\$92,446	\$91,761	2%
Canada	\$93,898	\$88,251	\$88,599	16%

\*2013 data

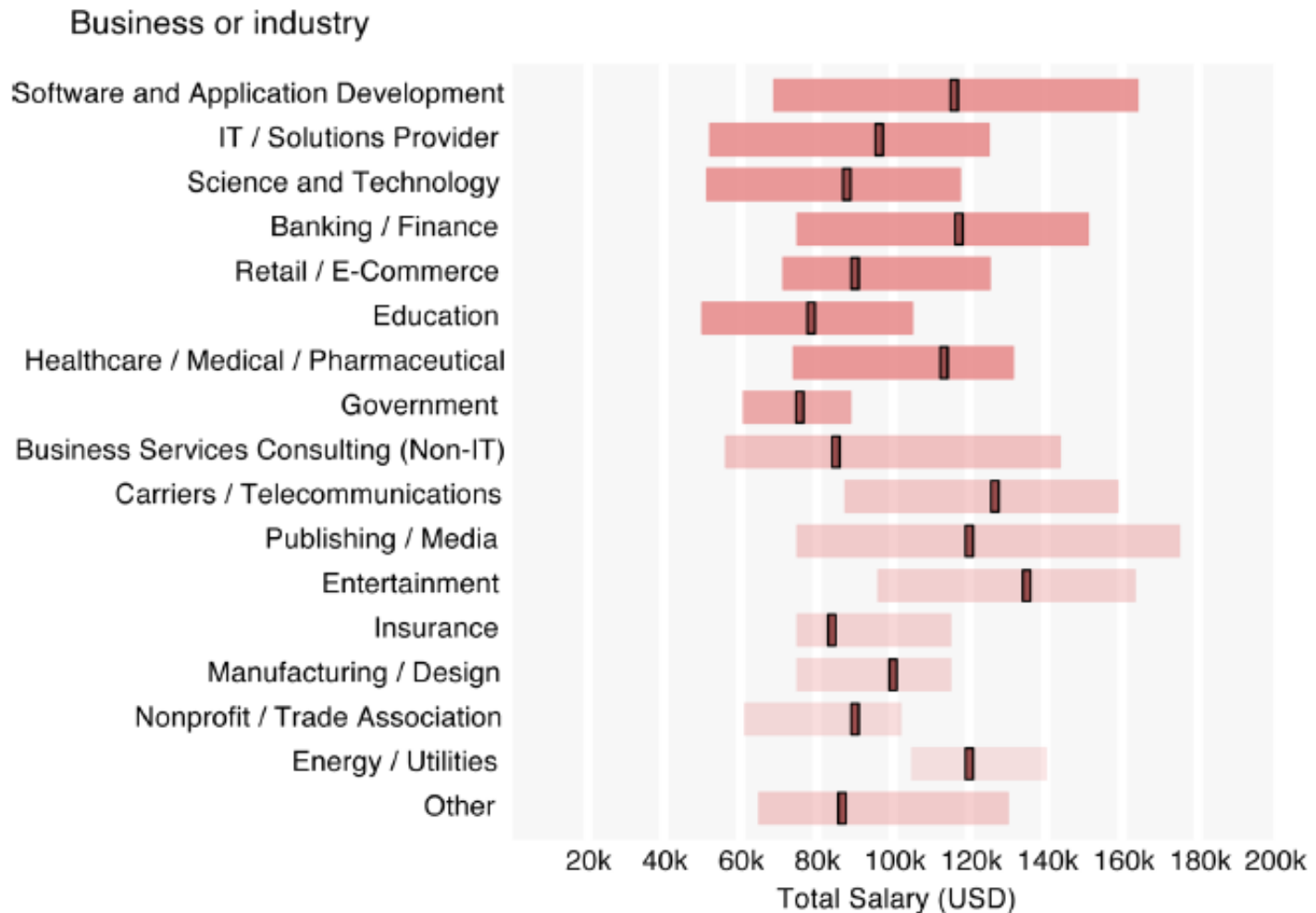
## AVERAGE SALARY BY INDUSTRY

	2011	2012	2013	Change*	Respondents**
Media/entertainment/publishing	\$116,363	\$110,357	\$129,114	+17.0%	2%
Consulting/professional services	\$115,237	\$120,395	\$115,726	-3.9%	13%
Software/Internet	\$112,238	\$113,953	\$110,791	-2.8%	7%
Manufacturing (computers and non-computers)	\$97,276	\$109,223	\$109,995	<1.0%	6%
Food/beverage	\$118,738	\$103,300	\$109,989	+6.5%	2%
Hospitality/travel	\$113,213	\$111,534	\$108,367	-2.8%	2%
Government (federal)	\$102,229	\$119,215	\$108,178	-9.3%	2%
Financial services	\$113,076	\$109,056	\$106,942	-1.9%	14%
Retail/wholesale/distribution	\$112,337	\$106,592	\$106,545	<1.0%	5%
Telecommunications	\$98,880	\$101,531	\$106,006	+4.4%	3%
Healthcare	\$99,633	\$102,412	\$105,034	+2.6%	13%
Insurance	\$99,333	\$100,345	\$104,544	+4.2%	10%
Education	\$85,113	\$92,768	\$95,170	+2.6%	5%
Transportation/logistics	\$99,180	\$102,531	\$92,672	-9.6%	3%
Government (state/local)	\$89,201	\$87,173	\$83,695	-4.0%	3%

\*Y-Y 2012-13

\*\*2013 data. Column does not total 100% because industries with lower representation were excluded.

# Salaries by Industry



Source: 2014 Data Science Salary Survey. O'Reilly Media, Inc. CA. Retrieved from <http://www.oreilly.com/data/free/files/2014-data-science-salary-survey.pdf>