

# Using Data Science to Recruit Data Scientists

Maryam Jahanshahi Ph.D.

Research Scientist

TapRecruit.co

**tap**Recruit.co

<http://bit.ly/pydata-dc-2018>

# The War on Talent is Real



**I Am Recruitr**  
@iamrecruitr

Following



## JOB AD:

- \* Must have 10+ years Swift experience
- \* Must have a PhD in Aqueous Geochemistry from a top tier University
- \* Must be a team player

\*zero applications later\*

Founder Blog Post:  
The War For Talent is real

5:07 AM - 6 Jul 2018



**I Am Devloper**  
@iamdevloper



> we're looking for a junior developer with the experience of a senior developer for the salary of an intern

12:10 PM · Oct 18, 2017 · Twitter Lite

**8.3K** Retweets   **17.2K** Likes

We know that what you  
say to candidates matters

# Research at TapRecruit

Helping companies make better recruiting decisions

## Decision Science:

- How do candidates make decisions about which jobs to apply to?

*$Pr(\text{apply}) = jd + culture + co + other\ jobs...$*

- How do hiring teams make decisions about candidate qualifications?

*$Pr(\text{success}) = candidate + jd + other\ candid...$*

## NLP and Data Science:

- What are distinguishing characteristics of successful career documents?

- What skills are increasingly important for different industries?

# TapRecruit uses NLP to understand career content

Converting unstructured documents into structured data



## **Smart Editor for JDs**

Data-driven suggestions on both the content and language use in job descriptions.



## **Pipeline Health Monitoring**

Analytics dashboards to help diagnose quality and diversity issues in talent pipelines.



## **Salary Estimation**

Data-driven salary estimates based on a job's requirements rather than just title and location.



Job ▾

Sync

Similar Jobs ▾

Open

Large Candidate Pool

Applicants: 202 ▾

3850 Characters

Notify ▾

Last edit: System ▾

28

Job will perform poorly

This job scores **lower than 95%** of **Junior Accounting** jobs in **Los Angeles, CA**

- Add preferred qualifications
- Add more "you" statements
- Perks included
- Equal opportunity statement is included

Neutral

Gendered



## Senior Finance

TapRecruit - Los Angeles



This job has not attracted enough qualified applicants.

Days Live:	50	Applied:	202
Screening:	14	Assessment:	1
Offer:	0	Hired:	0

[Job Report →](#)\$76,300 <sup>BETA</sup>

\$65,200

\$98,600

TapRecruit is looking for a smart, detail-oriented person to serve as a senior financial analyst. This person will be responsible for supporting the company's FP&A requirements. Responsibilities will include working on TapRecruit Entertainment Group's FP&A model, supporting analysis for long term planning, tracking key business operational metrics and producing monthly financial/operational reports. In addition to FP&A needs, this role will require strong organizational skills to help manage the department and evaluate/implement projects for top senior managers across the department and evaluate/implement projects for top management. This is a dynamic role that serves the finance department and will routinely interface with TapRecruit's top management. This is an ideal position for an individual who has gained strong experience at an investment bank or accounting firm and now seeks to apply those skills to a fast-growing entrepreneurial company. Strong quantitative and excel financial modeling skills are a must. The ideal candidate must be comfortable in a dynamic start-up environment, will bring energy and passion to everything he/she does, and will not be afraid to roll up his/her sleeves to tackle challenging analytical assignments.

Language that emphasizes an "intense" or "confusing" environment is known to deter qualified candidates.

[Delete](#)

This job is full-time, based in Los Angeles. We offer competitive compensation and stock option program.

# Language matters in job descriptions

Same title,  
Different job

<b>Finance Manager</b> <b>Kraft Foods</b>	<b>Finance Manager</b> <b>Roche</b>
Junior (3 Years)	Senior (6-8 Years)
No Managerial Experience	Division Level Controller
	Strategic Finance Role
	MBA / CPA

- ✓ Same Title
- ✗ Required Experience
- ✗ Required Responsibility
- ✗ Preferred Skill
- ✗ Required Education

Different title,  
Same job

<b>Performance Marketing Manager</b> <b>PocketGems</b>	<b>Senior Analyst, Customer Strategy</b> <b>The Gap</b>
Mid-Level	Mid-Level
Quantitative Focus	Quantitative Focus
iBanking Expertise	Finance Expertise
Data Analysis Tools (SQL)	Relational Database Experience
Consulting Experience Preferred	External Consulting Experience Preferred
MBA Preferred	BA in Accounting, Finance, MBA Preferred

- ✓ Required Experience
- ✓ Required Skills
- ✓ Required Experience
- ✓ Required Skills
- ✓ Preferred Experience
- ✓ Preferred Education

# Understanding the 'core' of a job

## Data

**15 M+**

Job Descriptions

**2.5 M+**

Verified Salaries

Real-time Candidate  
Job Search Data

## Classification

**500+**

Job Categories

**10,000+**

Unique Characteristics

Responsibilities  
Skills  
Education

## Guidance

**57,000+**

Unique Highlights

**500+**

Rule Kinds

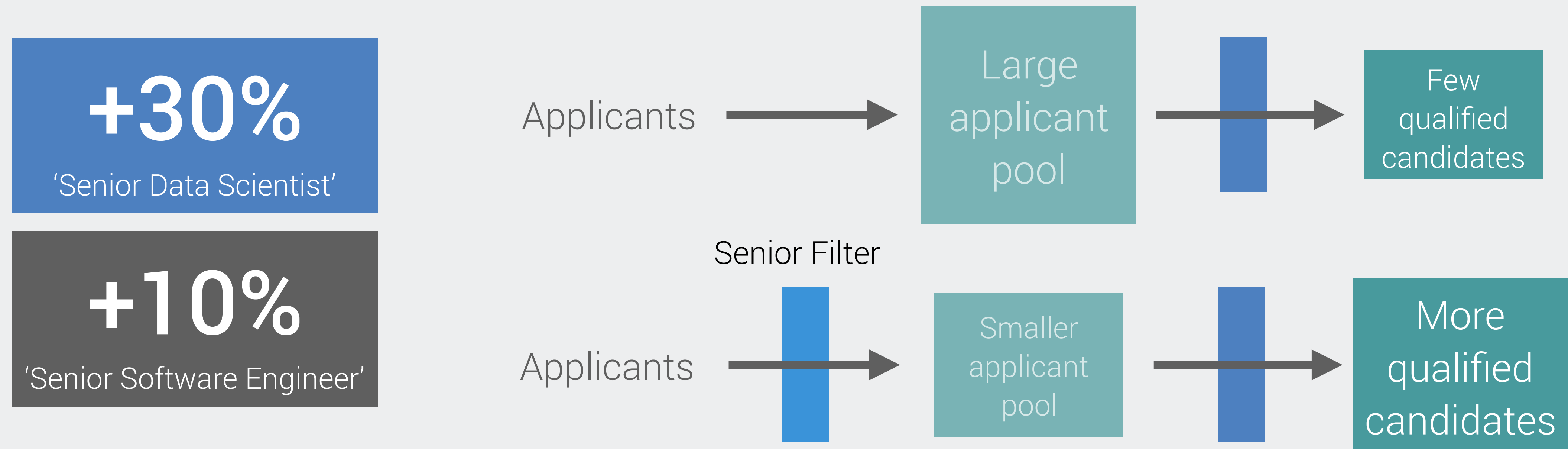


HEURISTIC 1:

**Inflating job titles improves  
applicant pool quality**

# The growth and impact of inflated titles

Hypothesis: Inflated titles amplify signal and reduce noise



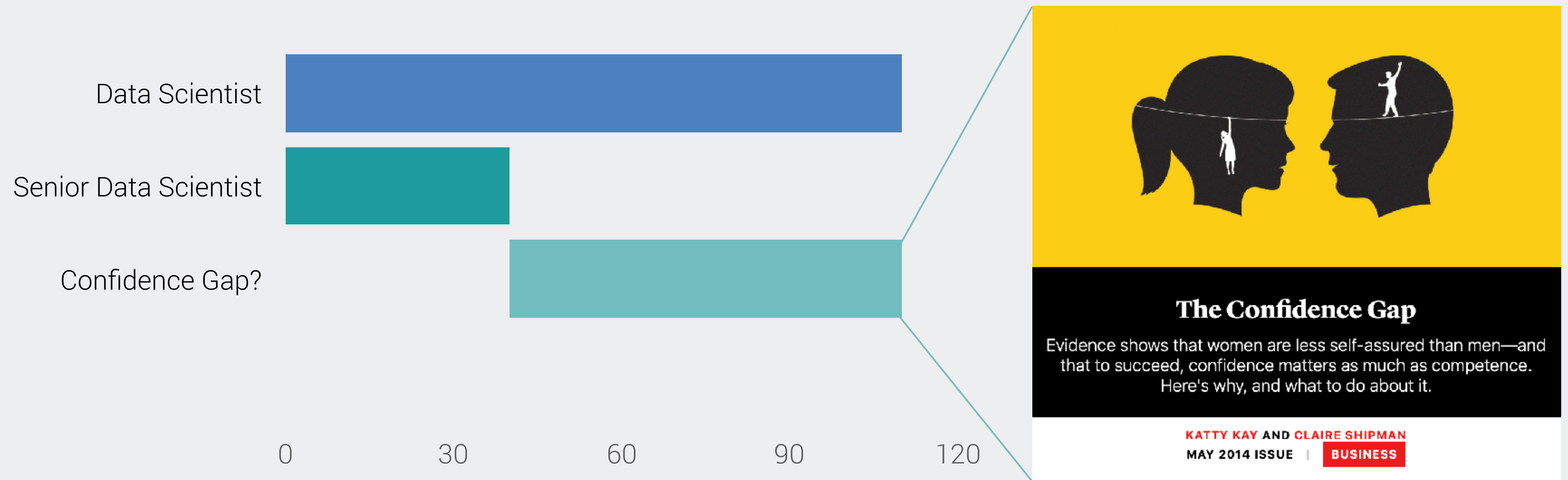
# Title inflation does not live up to its promise

Fewer qualified applicants, lower chance of a successful hiring process

	Senior Data Scientist	Data Scientist	
Average Total Applicants	40	110	
Average Qualified Candidates	4	8	2x
Chance of Success	29%	58%	

# The impact of one word

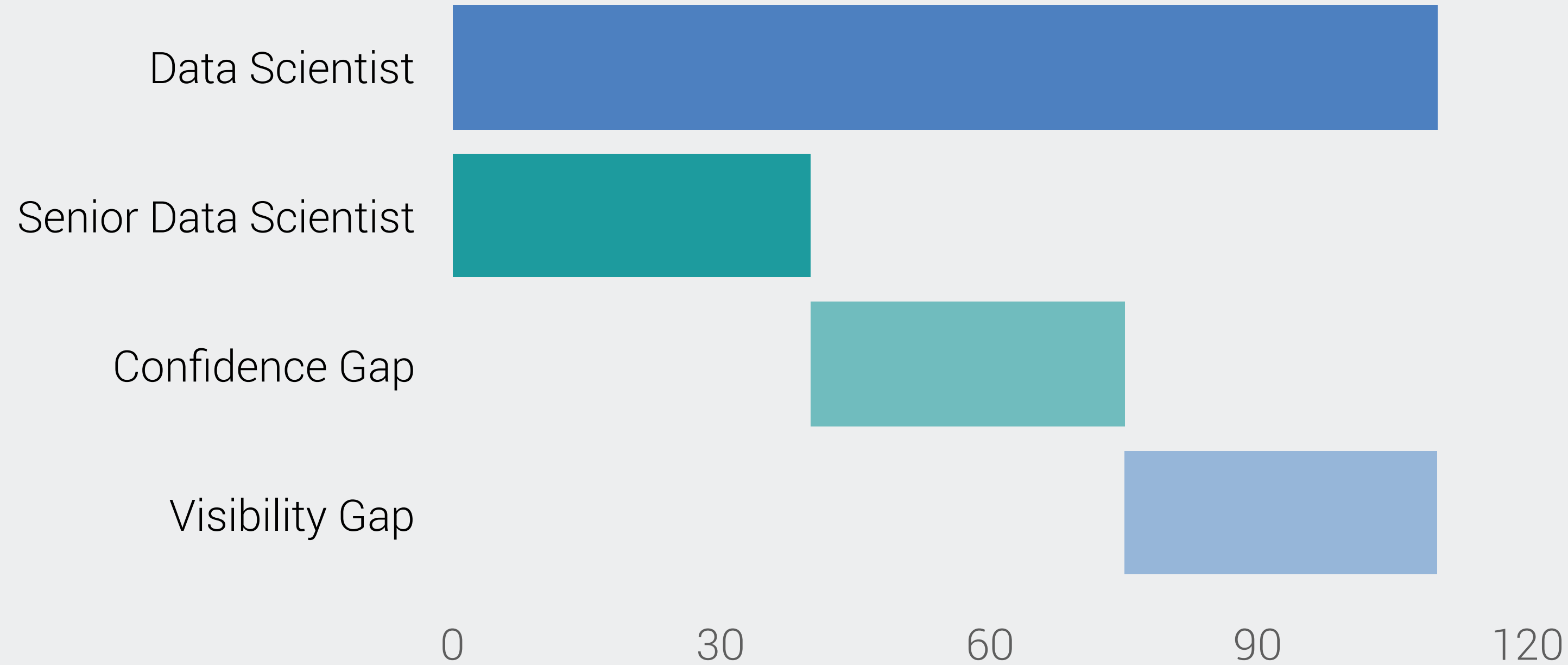
What accounts for the difference in candidate pool size?





# Jobs with inflated titles are hit twice

Inflated titles drive fewer job views and fewer applications



**Rare Search Term:** Candidates rarely search for 'Senior Data Scientist' roles

**Search 101:** When candidates search for 'Data Scientist' roles, 'Senior Data Scientist' roles are considered weaker results by search algos and less frequently surfaced.

# Jobs with inflated titles attract fewer women

Fewer female candidates and a lower chance of a female hire

	Senior Data Scientist	Data Scientist	
Female Candidates	8 (23%)	39 (36%)	4x (+50%)
Qualified Female Candidates	1 (26%)	2 (32%)	2x (+23%)
Chance of Female Hire	15%	39%	

HEURISTIC 2:

**It's all about the money**

# How candidates can read between lines



**I Am Recruit**  
@iamrecruit


Following

▼

SALARY: \$40k to \$60k

TRANSLATION: \$40k

9:59 AM - 6 Aug 2018



**I Am Recruit**  
@iamrecruit

Following

▼

PERKS: A competitive salary, a ping-pong table, beers on a Friday!

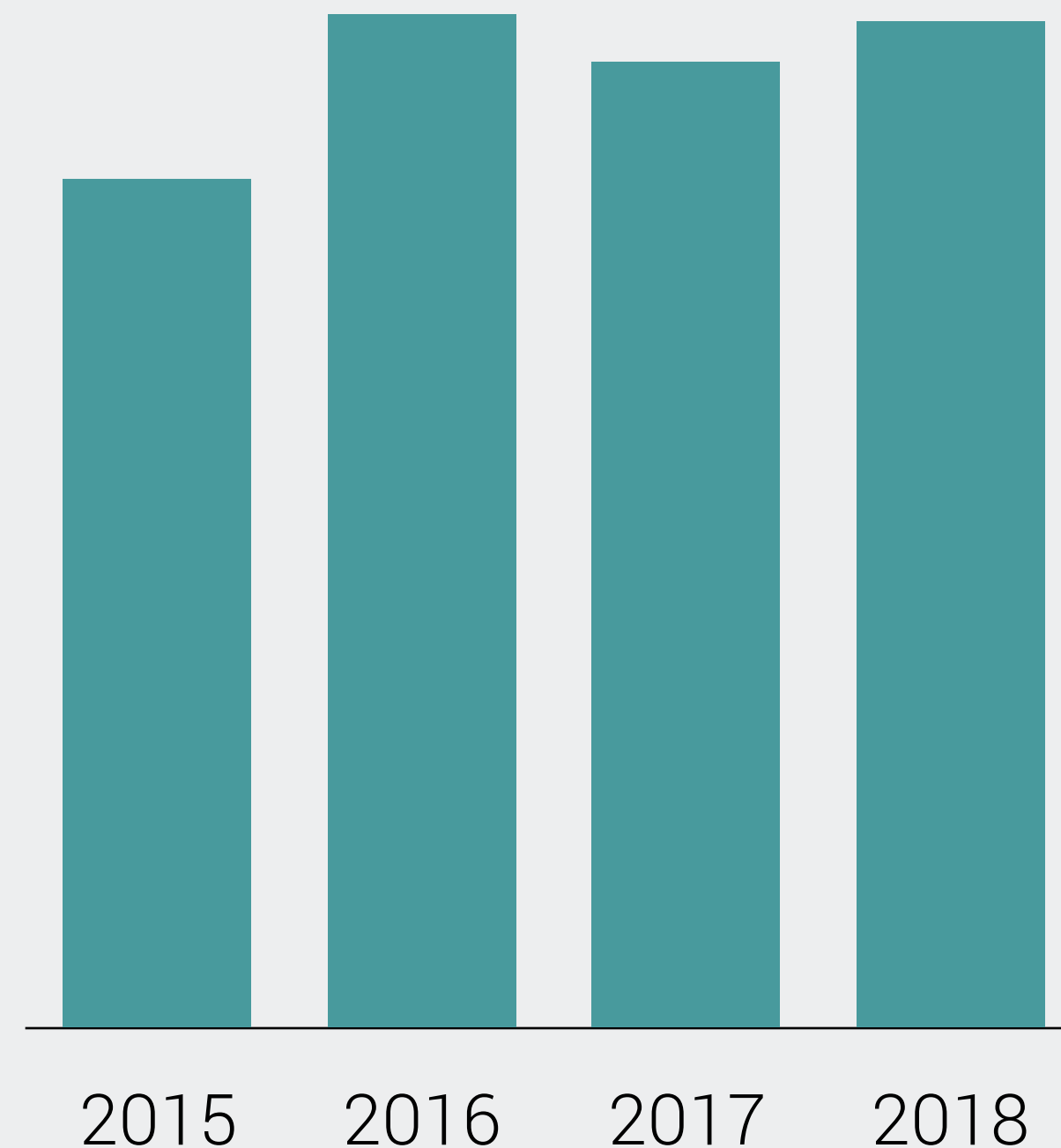
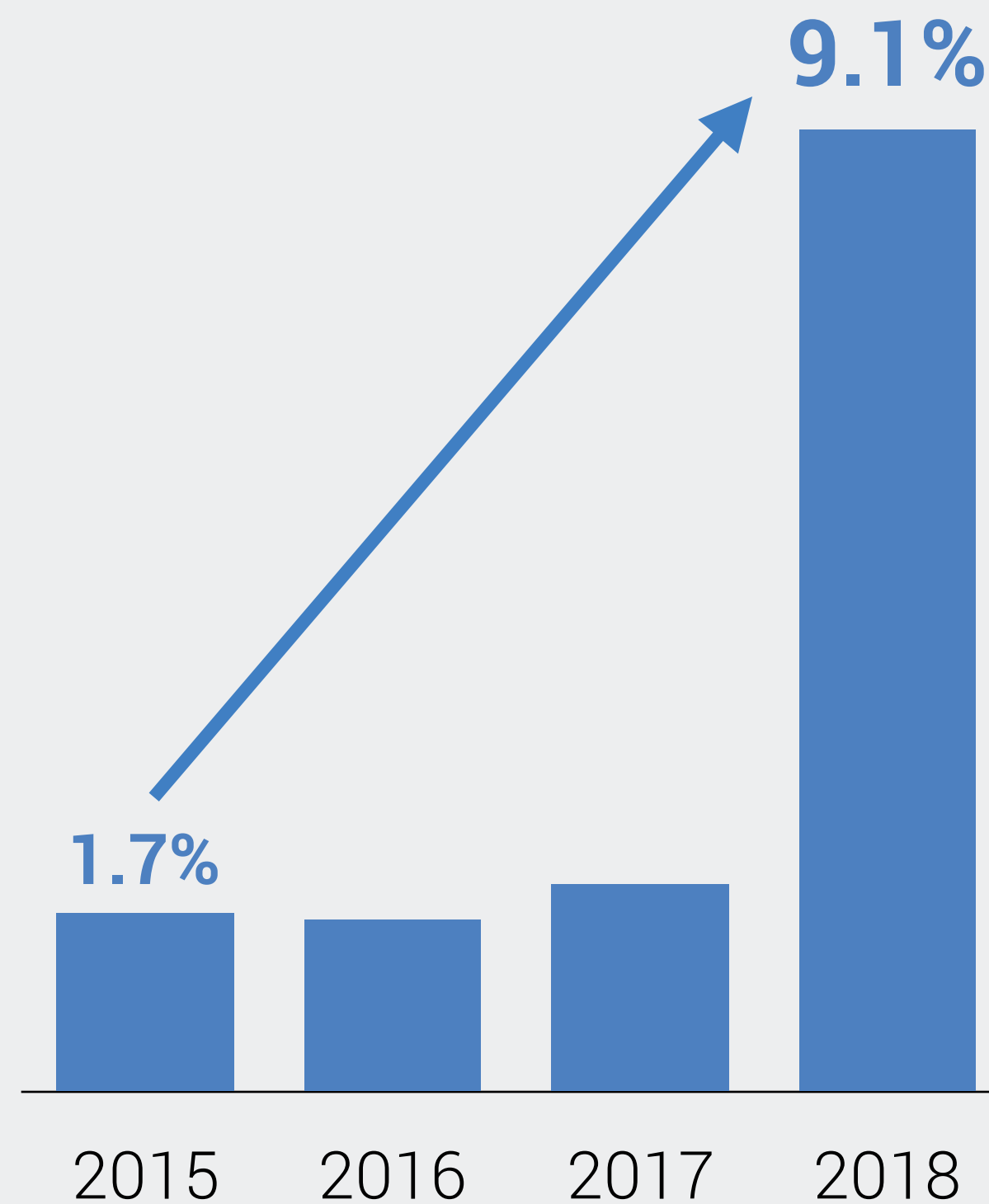
TRANSLATION: We have no idea what we're doing

6:06 AM - 18 Sep 2018



# The paradox of the 'competitive salary'

Dramatic rise in JDs, while real salaries have remained steady



Jobs that mention  
'competitive salary'

**Attract 20% fewer  
candidates**

**Take 10-15%  
longer to fill**

# 62% of jobseekers consider perks and benefits critical factors in deciding on jobs

but few companies mention their perks and benefits

F500 + Public Co.

16%

Unicorns

18%

Startups

40%

# Do unusual P&B attract candidates?

Standard P&B resonate but phrasing matters

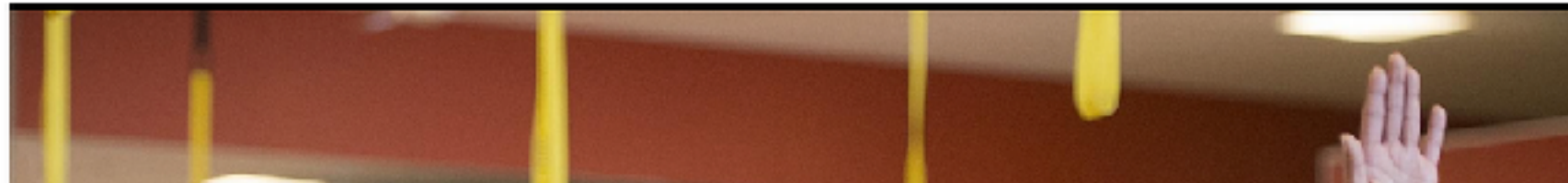
## BUSINESS

### Happy Workers, Richer Companies?

There is a free lunch in economics, after all. Companies with perks like complimentary meals clearly outperform competitors in the stock market.

DEREK THOMPSON JUL 28, 2014

### Companies offering unique perks to attract millennials



## FINDING A JOB

### 13 Companies as Amazing as Apple (and Hiring Now!)

Medical Insurance

401k

Family Leave

Paid Time Off

Unlimited Vacation

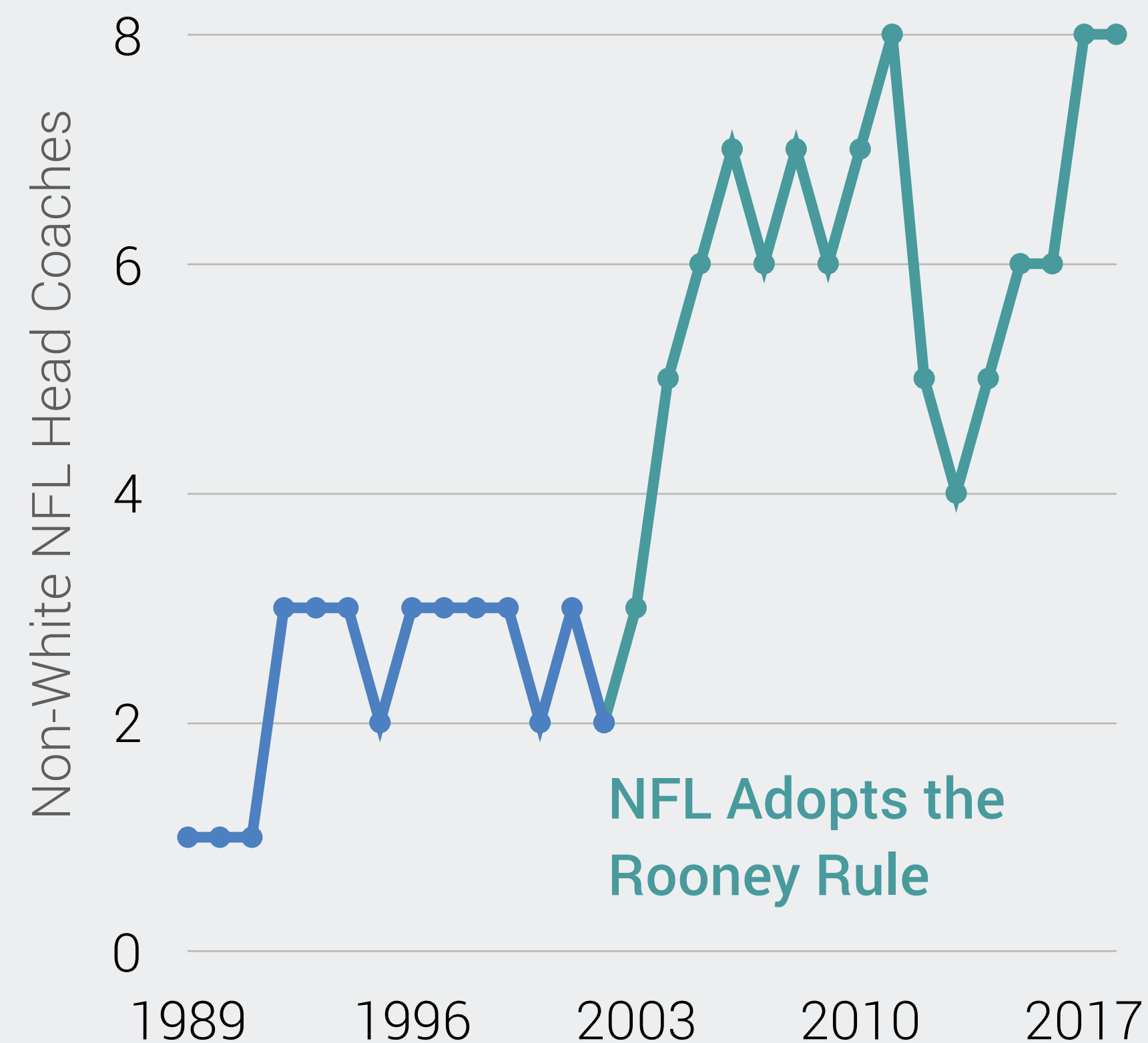
HEURISTIC 3:

**Diversity can be 'hacked'**



# The Rooney Rule combats lack of representation

Significant impact in the NFL (2003), Adoption in Big Tech (2016)



**Facebook is using the same approach the NFL took to increase diversity in the league**

Richard Feloni Jan. 24, 2016, 10:35 AM



SILICON VALLEY

**After Amazon Adopts Rooney Rule, Black Lawmakers Call for More Diversity in Tech**

Tech Accountability Caucus co-chair says congressional hearings, meetings can help keep pressure on Silicon Valley

PINTEREST

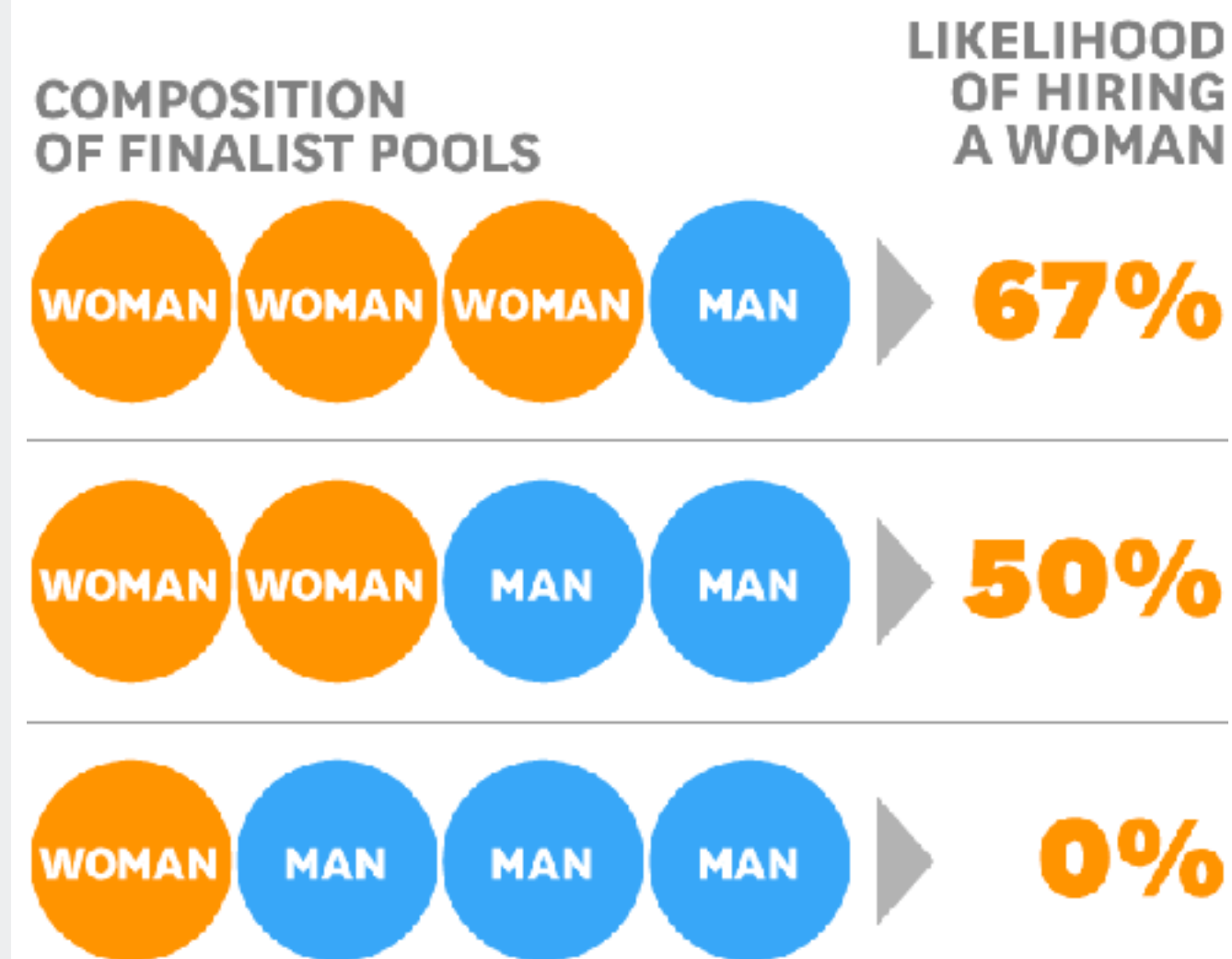
**What Pinterest is learning from the Pittsburgh Steelers about diversity**

# Real world implications of the Rooney Rule

## Efficacy in Actual Hiring Decisions, Candidate Backlash

### The Relationship Between Finalist Pools and Actual Hiring Decisions

According to one study of 598 finalists for university teaching positions.



SOURCE STEFANIE K. JOHNSON ET AL

© HBR.ORG



Eddie F of the Boyz  
@Aisha\_CG

Hiring managers and search committees: stop asking POC to apply to your jobs so you can have a diverse pool IF you already have folks in mind for the job. We DO NOT exist to make your search viable. We want to work, get paid, and support users in real and substantial ways

2:34 PM · Jul 8, 2018



EricaJoy  
@EricaJoy

Speaking from the perspective of a hiring manager: 100  
Speaking from the perspective of a Black woman who has better things to do with her time than be the Rooney Rule candidate in a sham of a recruiting process: 100 100 100

# Large candidate pools promote representation

	40 Candidates	100 Candidates	
Female Candidates	11 (27%)	38 (38%)	3x (+40%)
Qualified Female Candidates	1 (16%)	4 (34%)	4x (+100%)
<b>Chance of Female Hire</b>	<b>19%</b>	<b>43%</b>	



HEURISTIC 1:

## **Job title inflation improves applicant pools**

- Title inflation decreases signal with no change to noise
- The double-whammy of the visibility and confidence gaps

HEURISTIC 2:

## **It's all about the money**

- Competitive salary makes you uncompetitive
- The 'boring' benefits are not boring to candidates

HEURISTIC 3:

## **Diversity can be hacked**

- Representative candidate pools -> representative workforces



# Research at TapRecruit

Helping companies make better recruiting decisions

## Decision Science:

- How do candidates make decisions about which jobs to apply to?

*$Pr(\text{apply}) = jd + company + culture + \dots$*

- How do hiring teams make decisions about candidate qualifications?

*$Pr(\text{success}) = candidate + jd + other\ candid\dots$*

## NLP and Data Science:

- What are distinguishing characteristics of successful career documents?
- What skills are increasingly important for different industries?

How have data science skills  
changed over time?

# Word embeddings capture semantic similarities

Statistical modeling through software (e.g. SPSS) or programming language (e.g. **Python**)

**Context**

**Word**

Experience in **Python**, Java or other object-oriented programming languages

**Context**

**Word**

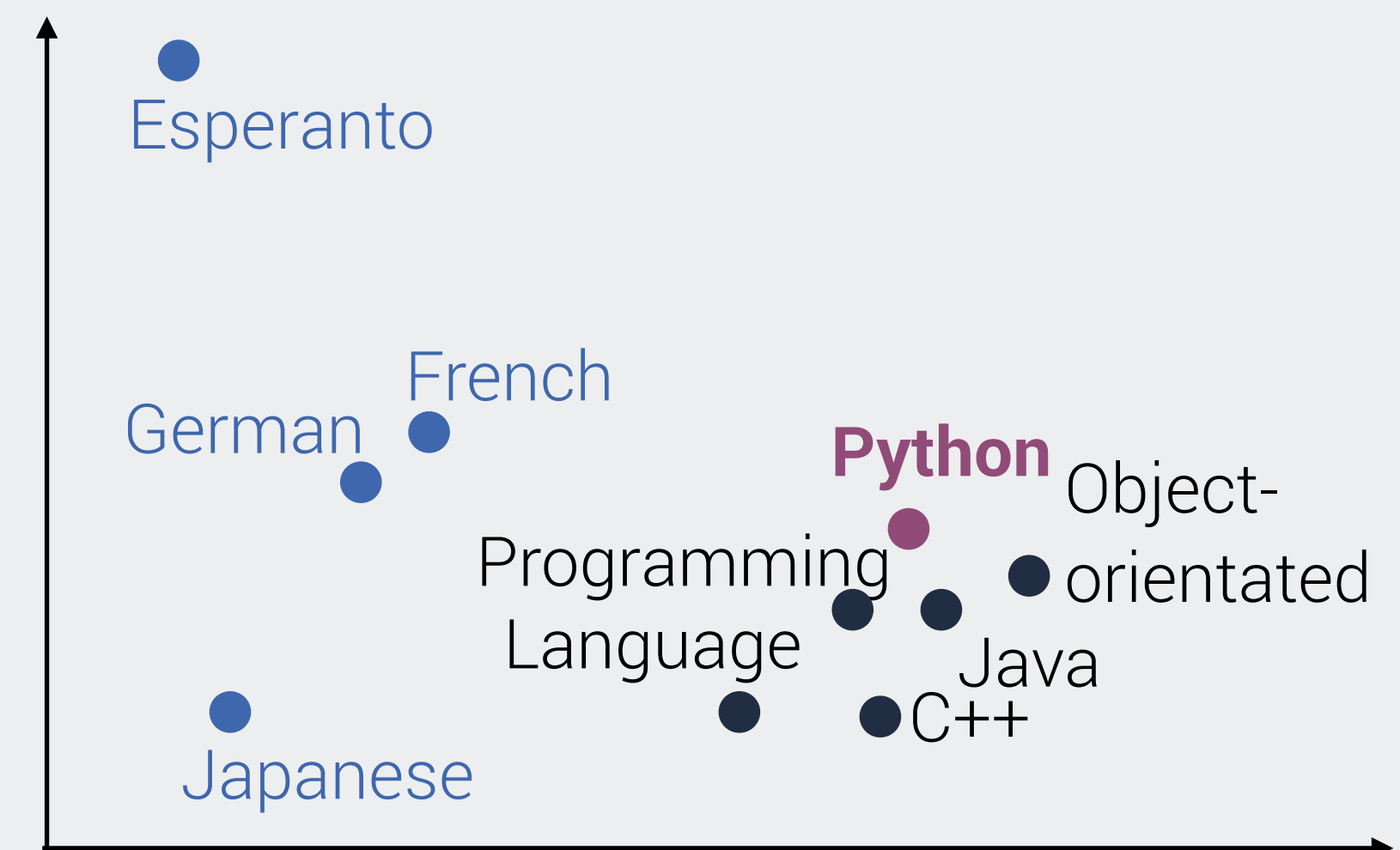
**Context**

Proficiency programming in **Python**, Java or C++.

**Context**

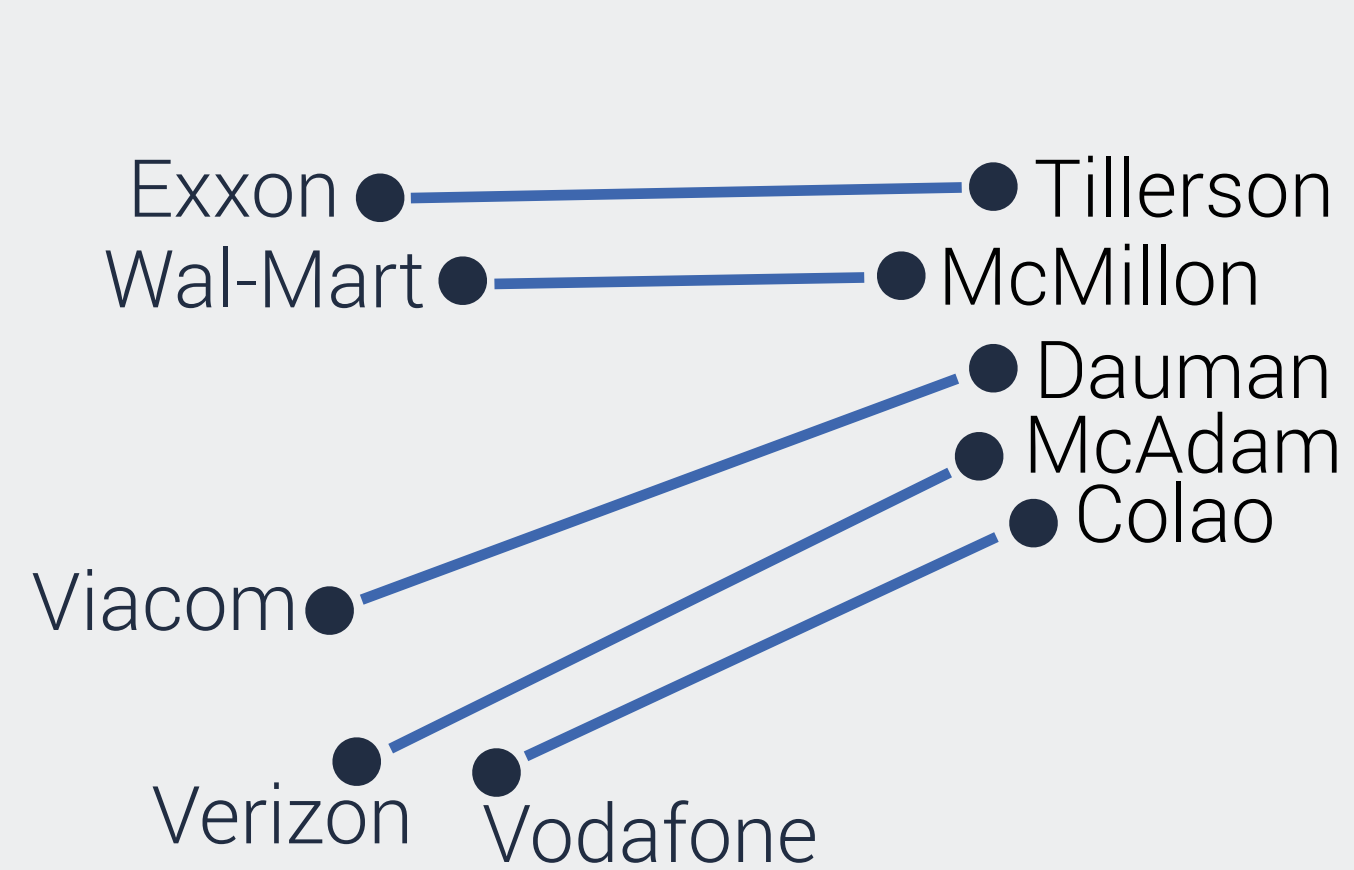
**Word**

**Context**

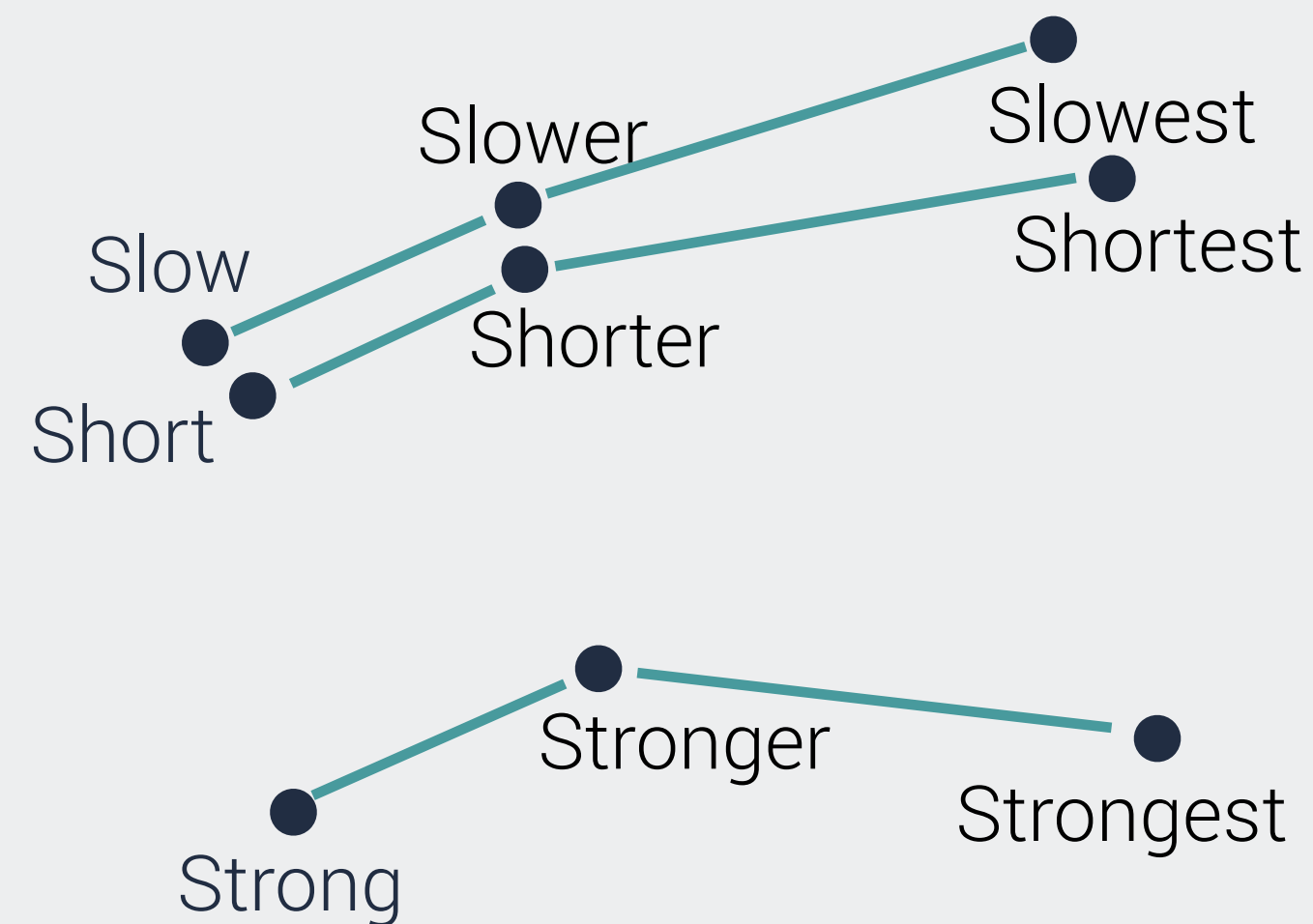


# Embeddings capture entity relationships

Dimensionality enables comparison between word pairs



**Hierarchies**



**Comparatives and Superlatives**

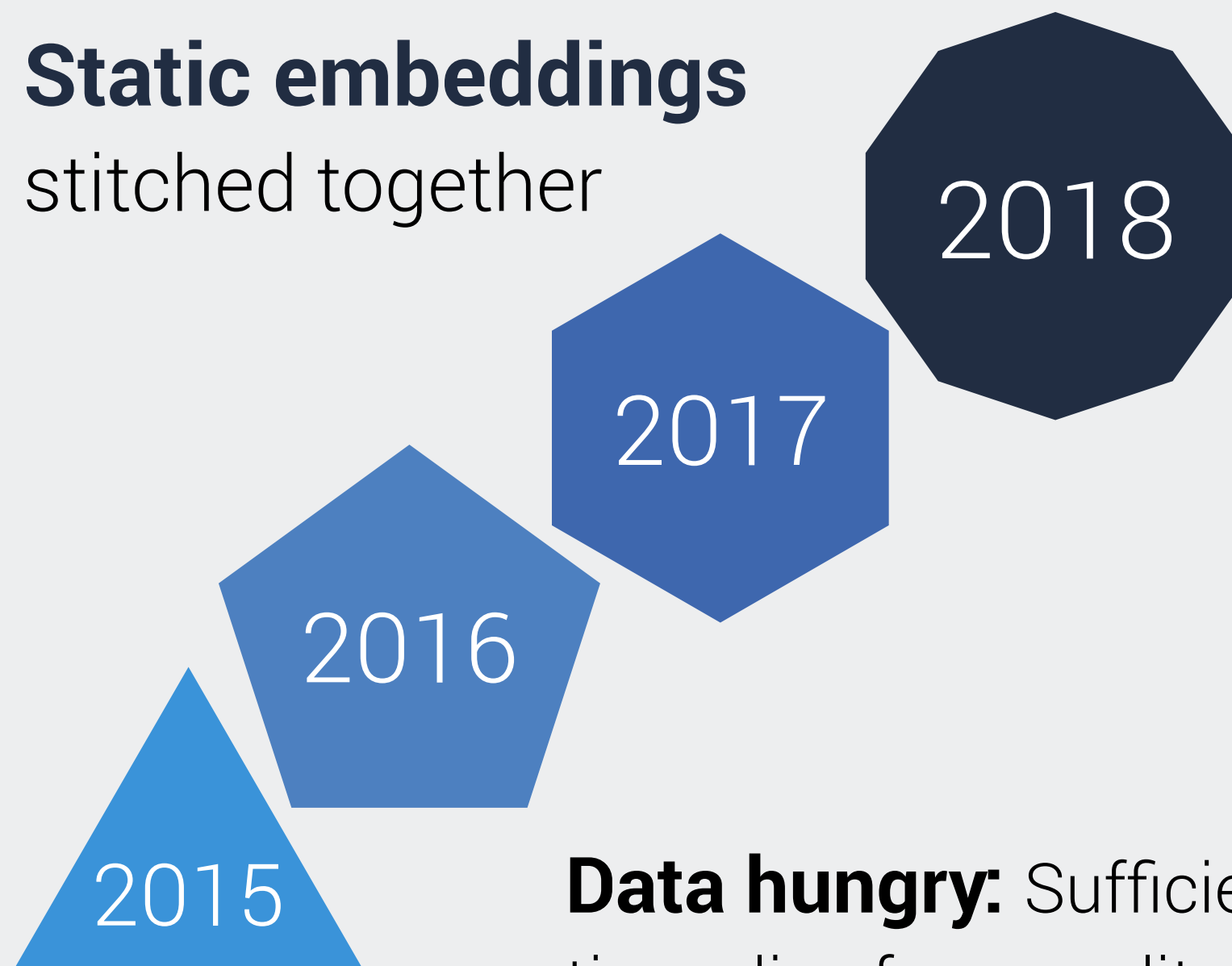


**Woman :: Queen as Man :: ?**

# Two approaches to connect embeddings

## Static embeddings

stitched together



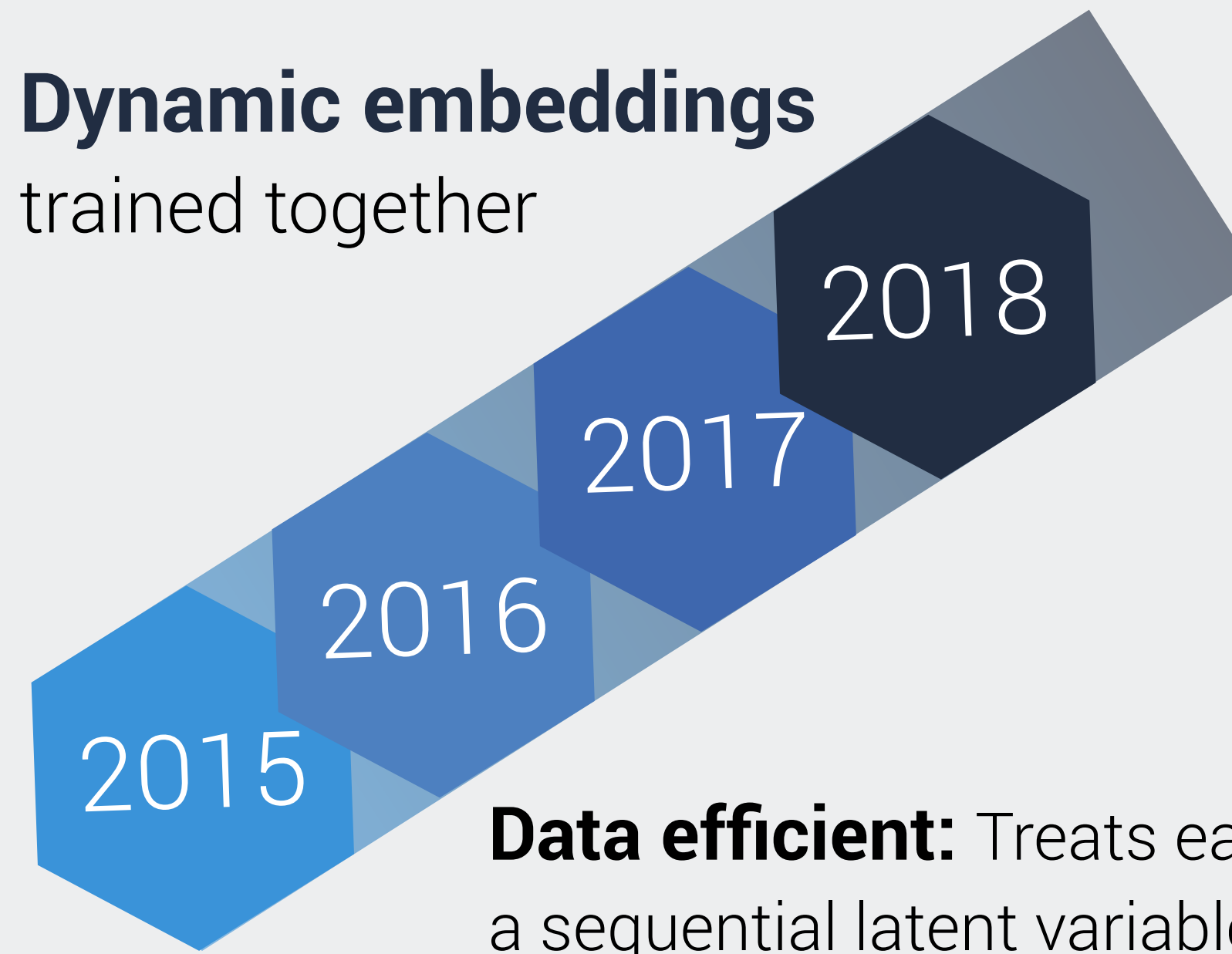
**Data hungry:** Sufficient data for each time slice for a quality embedding.

**Requires alignment:** Each time slice is trained independently. Dimensions are not comparable across slices.

Kim, Chiu, Kaneki, Hedge and Petrov, [arXiv: 1405:3515](#).  
Kulkarni, Al-Rfou, Perozzi and Skiena, [arXiv: 1411:3315](#).

## Dynamic embeddings

trained together



**Data efficient:** Treats each time slice as a sequential latent variable, enabling time slices with sparse data.

**Does not require alignment:** Treating time slice as a variable ensures embeddings are connected across slices.

Balmer and Mandt, [arXiv: 1702:08359](#)  
Yao, Sun, Ding, Rao and Xiong, [arXiv: 1703:00607](#)  
Rudolph and Blei, [arXiv: 1703:08052](#)



# Dynamic Embeddings

Rudolph and Blei, [arXiv: 1703:08052](https://arxiv.org/abs/1703.08052)

## Absolute drift

Identifies top words whose usage changes over time course

words with largest drift (Senate)			
IRAQ	3.09	coin	2.39
tax cuts	2.84	social security	2.38
health care	2.62	FINE	2.38
energy	2.55	signal	2.38
medicare	2.55	program	2.36
DISCIPLINE	2.44	moves	2.35
text	2.41	credit	2.34
VALUES	2.40	UNEMPLOYMENT	2.34

## Embedding neighborhoods

Extract semantic changes by nearest neighbors of drifting words

UNEMPLOYMENT		
1858	1940	2000
unemployment	unemployment	unemployment
unemployed	unemployed	jobless
depression	depression	rate
acute	alleviating	depression
deplorable	destitution	forecasts
alleviating	acute	crate
destitution	reemployment	upward
urban	deplorable	lag
employment	employment	economists
distressing	distress	predict

Repository Link: [http://bit.ly/dyn\\_bern\\_emb](http://bit.ly/dyn_bern_emb)

# Experiments with Dynamic Bernoulli Embeddings

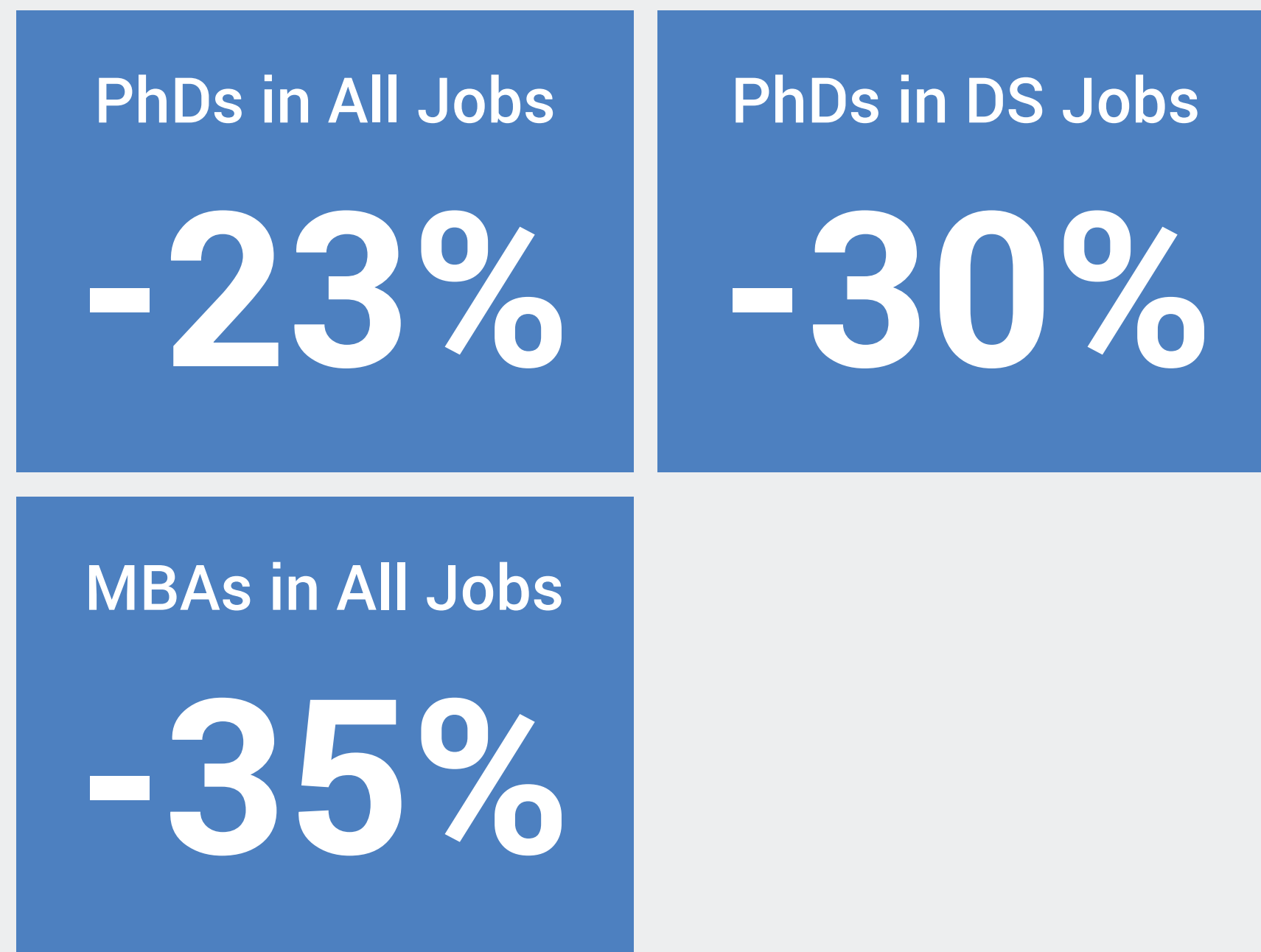
	Small Corpus	Large Corpus
<b>Job Types</b>	All	All
<b>Time Slices</b>	3 (2016-2018)	3 (2016-2018)
<b>Number of Documents</b>	<b>50 k</b>	<b>500 k</b>
<b>Embedding Training</b>	100 dimensions	100 dimensions

**Repository Link:** [http://bit.ly/dyn\\_bern\\_emb](http://bit.ly/dyn_bern_emb)

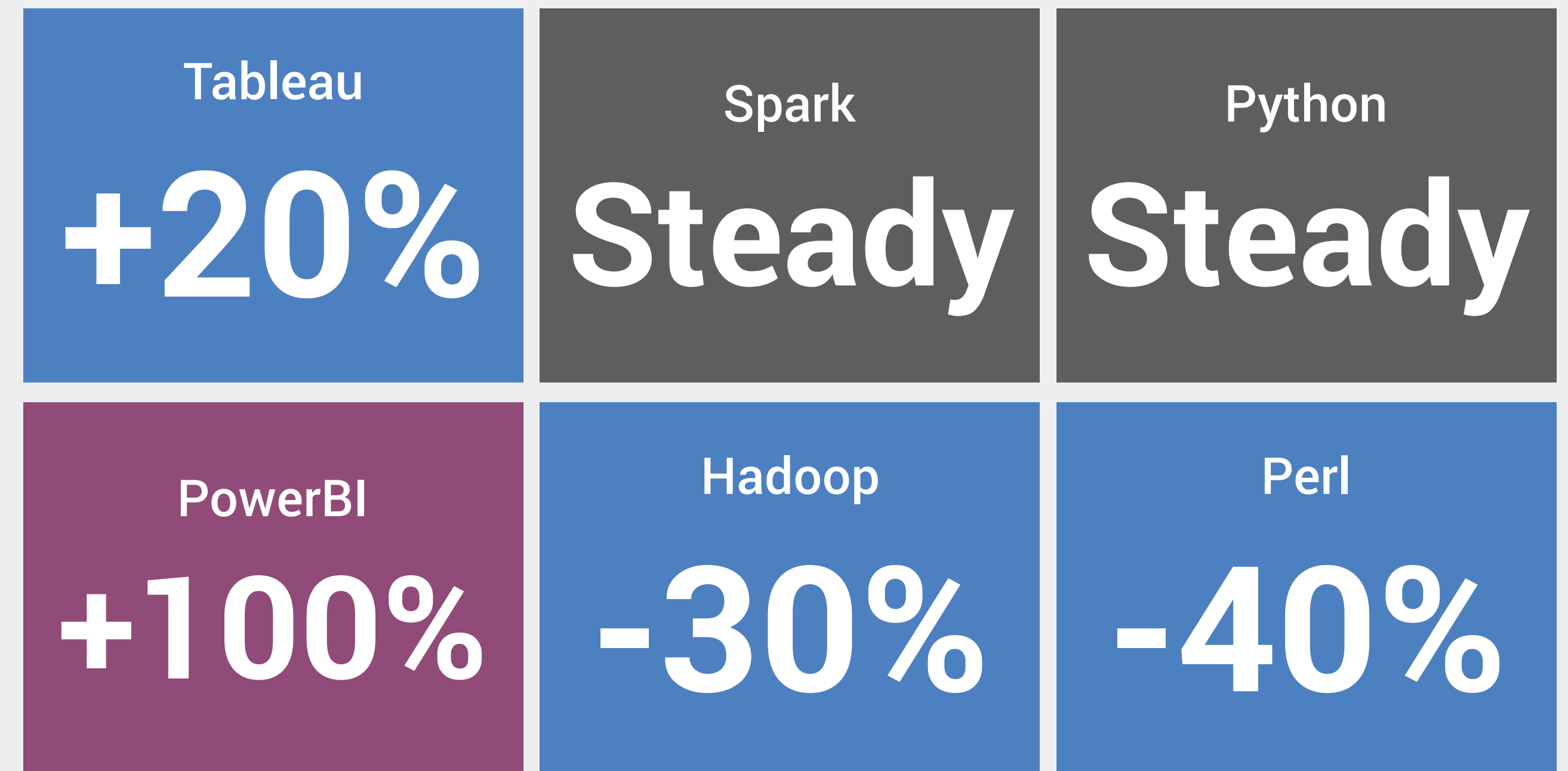
# Dynamic Bernoulli embeddings

Small corpus identified gains and losses

## Demand for PhDs and MBAs is Falling



## Data Science skills showing significant shifts



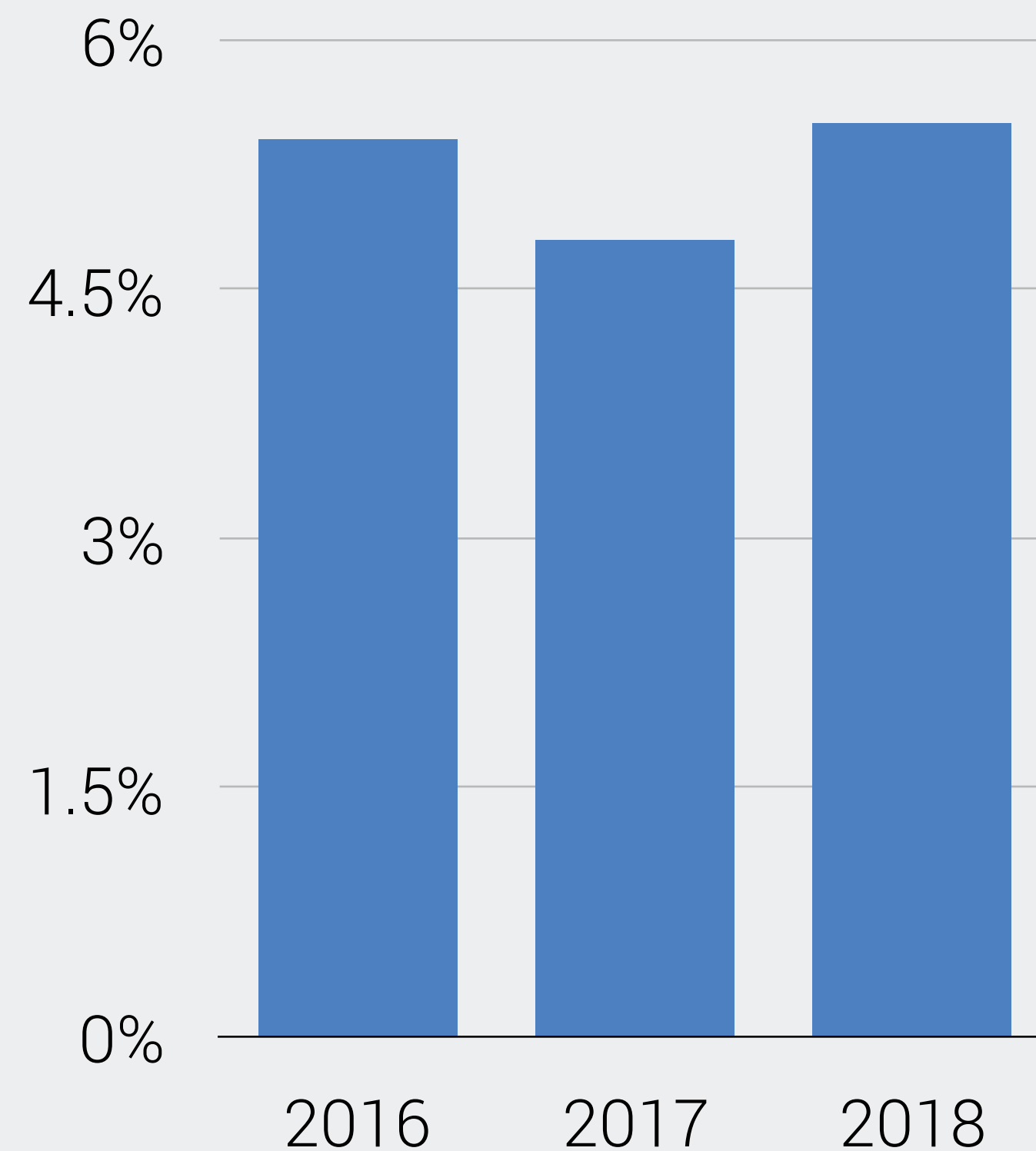
Blue boxes indicate phrases identified from top drifting words analysis.

Grey boxes indicate 'control' skills.

# Dynamic Bernoulli embeddings

Large corpus identified role-type dependent shifts in requirements

## No change to SQL demand



## SQL requirement increases in specific functions



# Thank you PyData DC!

**Maryam Jahanshahi Ph.D.**

Research Scientist

 @mjahanshahi

 maryam-j

**tap**Recruit.co

<http://bit.ly/pydata-dc-2018>