



Silicon Valley vs. New York:

who has better data scientist?

D.Vrdoljak vs. G.Kleemann PhD.



Who Has Better Data Scientists?

New York

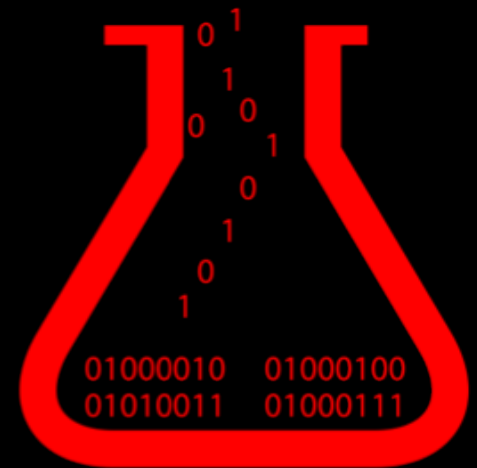
Silicon Valley



About us (BDSG)

Berkeley Data Science Group

Founded by UC Berkeley Data Science Instructors and Alumni with the goal of bringing MIDS work to market and commercializing Data Science research.



About us (the speakers)

Denis Vrdoljak
denis@bds.group

Gunnar Kleemann, PhD
gunnar@bds.group



About us (the speakers)

Denis Vrdoljak
denis@bds.group

Gunnar Kleemann, PhD
gunnar@bds.group

the Hippie

the Yankee



Who Has Better Data Scientists?

How Can We Answer This?



Some of our Related Tools and Projects

SciLit

MarketGraph

BioPredictor

Project “Pigeon”



We'll Use These to Answer the Question

SciLit

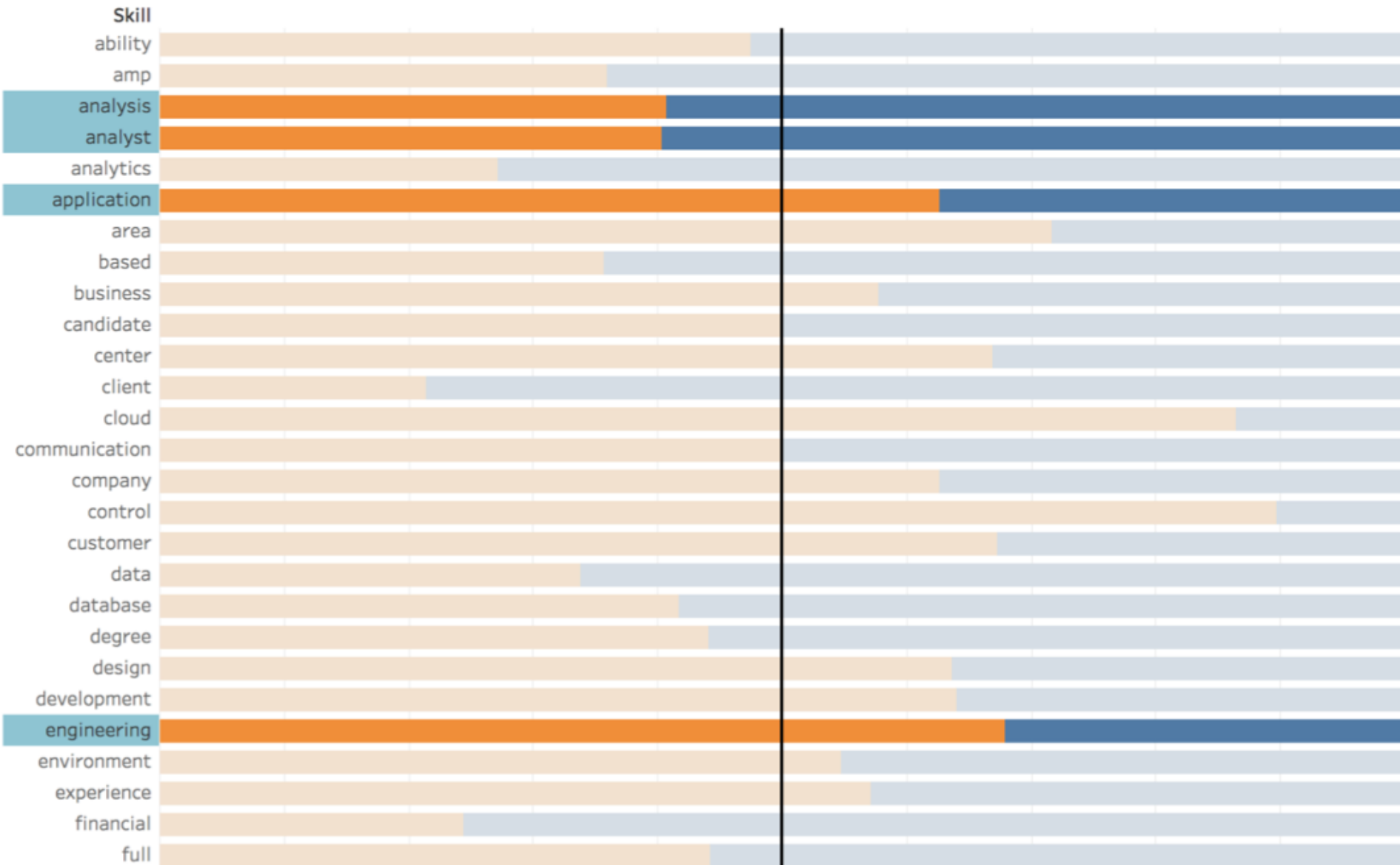
Project “Pigeon”



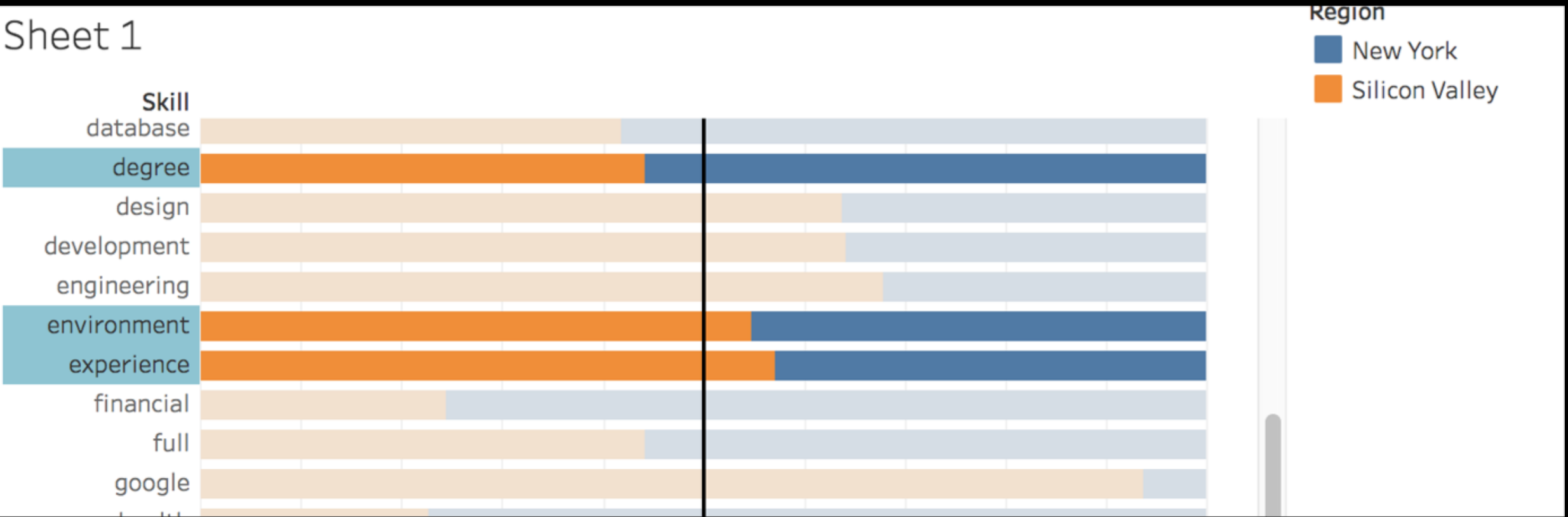
Sheet 1

Region

- New York
- Silicon Valley



Sheet 1



localhost

skillsAnalysis_DenisVersion PigeonAustin2018.pptx - Google Slides

Jupyter skillsAnalysis_DenisVersion Last Checkpoint: 10 hours ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Help Trusted Python 2

Code

```
    ('gender identity', 3),
    ('world class', 3)]

In [16]: relatedskills('data science','NY',5)

Out[16]: [('data science', 3885),
          ('big data', 616),
          ('data engineering', 312),
          ('science communication', 310),
          ('information system', 297)]

In [20]: relatedskills('Python','SV',5)

Out[20]: [('python', 328),
          ('similar', 137),
          ('unix', 136),
          ('programming experience', 21),
          ('language', 15)]

In [21]: relatedskills('Python','NY',5)

Out[21]: [('plus', 636),
          ('programming experience', 615),
          ('python', 383),
          ('sql', 324),
          ('r', 320)]

In [ ]: |
```



Project Pigeon

“Don’t get PIGEON holed by your resume.”



Project Pigeon:

Exchangeable relationships

- Which skills are Interchangeable or related?
 - R and Python?
 - C++ and JAVA?
- Which keyword exchanges or additions would increase matches?
- This applies to both resumes (Job Seekers) and job postings (Hiring Managers)



Project Pigeon

51,700

Job Skills on LinkedIn

20,000 (Jobs /major metro area /job title)

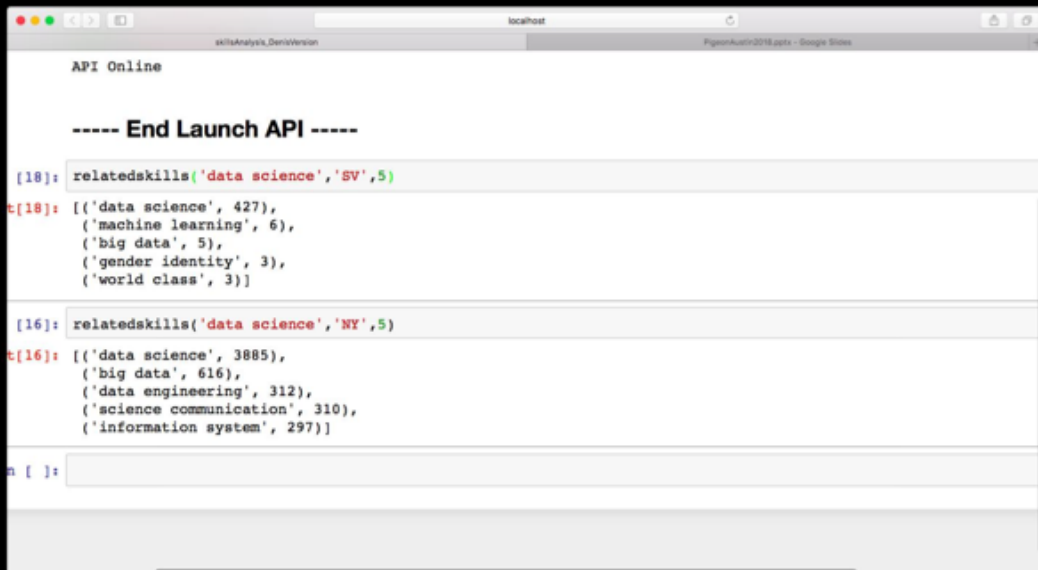
2,000,000

Relationships to score, per search, after cleaning data



Project Pigeon

Focus of Project Pigeon is the API and model customization



```
API Online

----- End Launch API -----

[18]: relatedskills('data science','SV',5)
t[18]: [{"data science": 427},
        {"machine learning": 6},
        {"big data": 5},
        {"gender identity": 3},
        {"world class": 3}]

[16]: relatedskills('data science','NY',5)
t[16]: [{"data science": 3885},
        {"big data": 616},
        {"data engineering": 312},
        {"science communication": 310},
        {"information system": 297}]

n [ ]:
```



Scilit

Using scientific literature to assess tech centers



Publication Data:

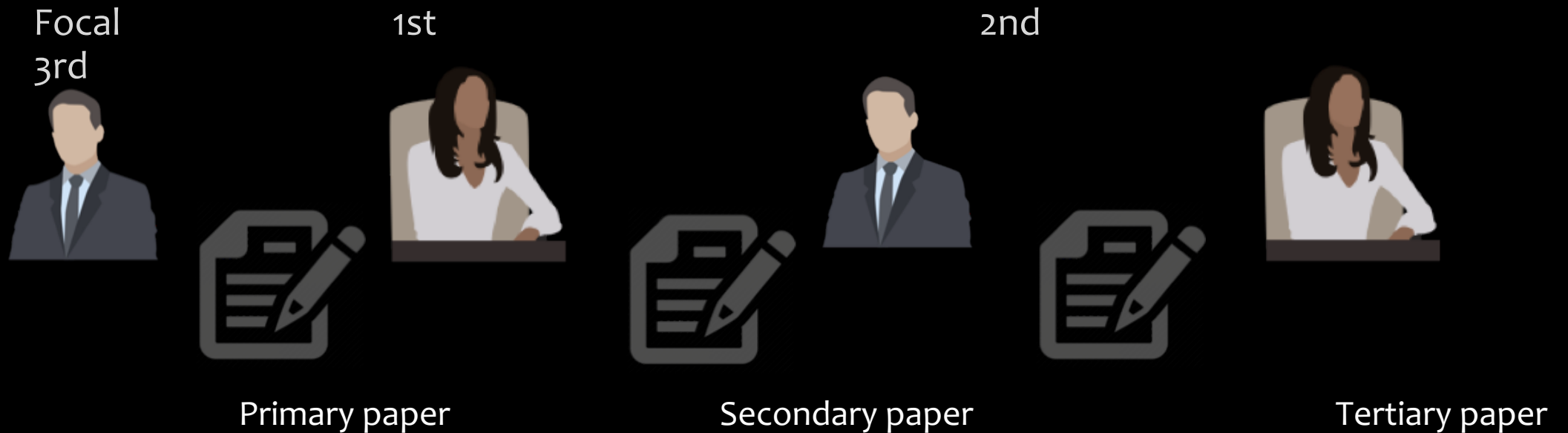
24,000,000
Full PubMed Dataset (all lit)

2,900,000 (~1.5 Million Open Source)
PubMed Central (BioTech -open access subset)

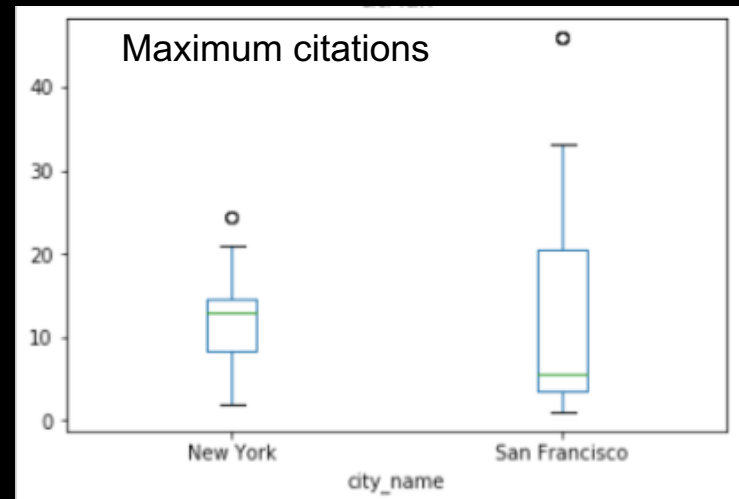
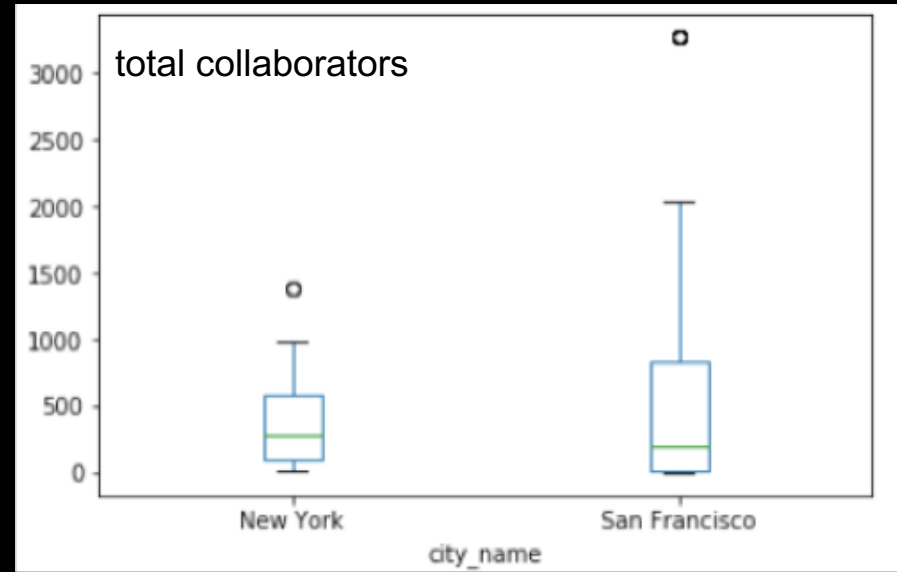
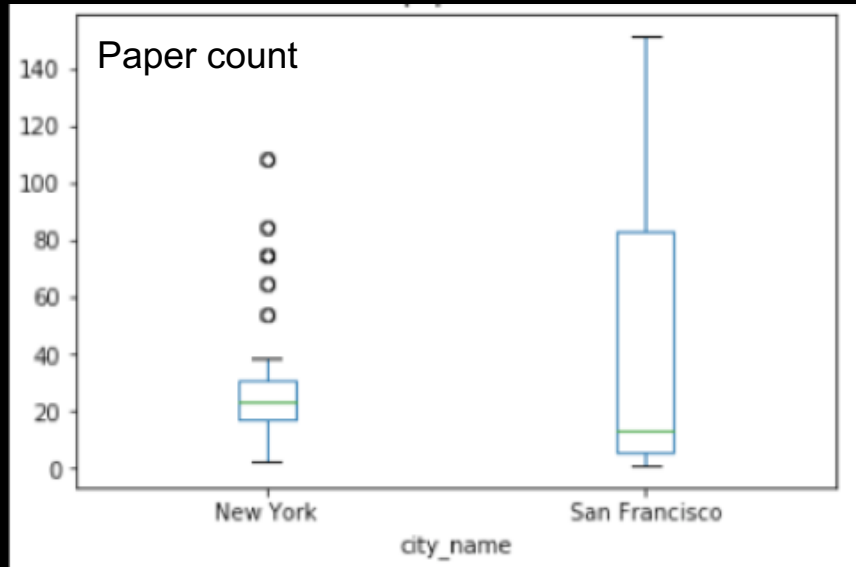
1,219,850
Total Entries, Condensed, after Cleaning/Pre-Processing



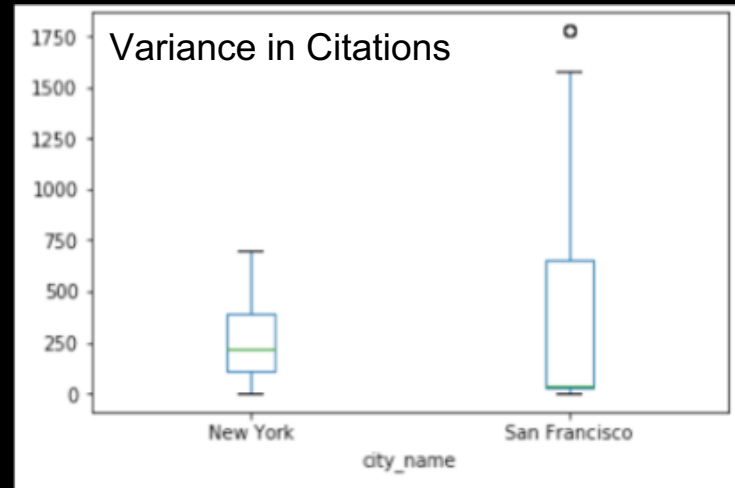
Building the collaboration graph in 3 layers



But how do we judge scientists?

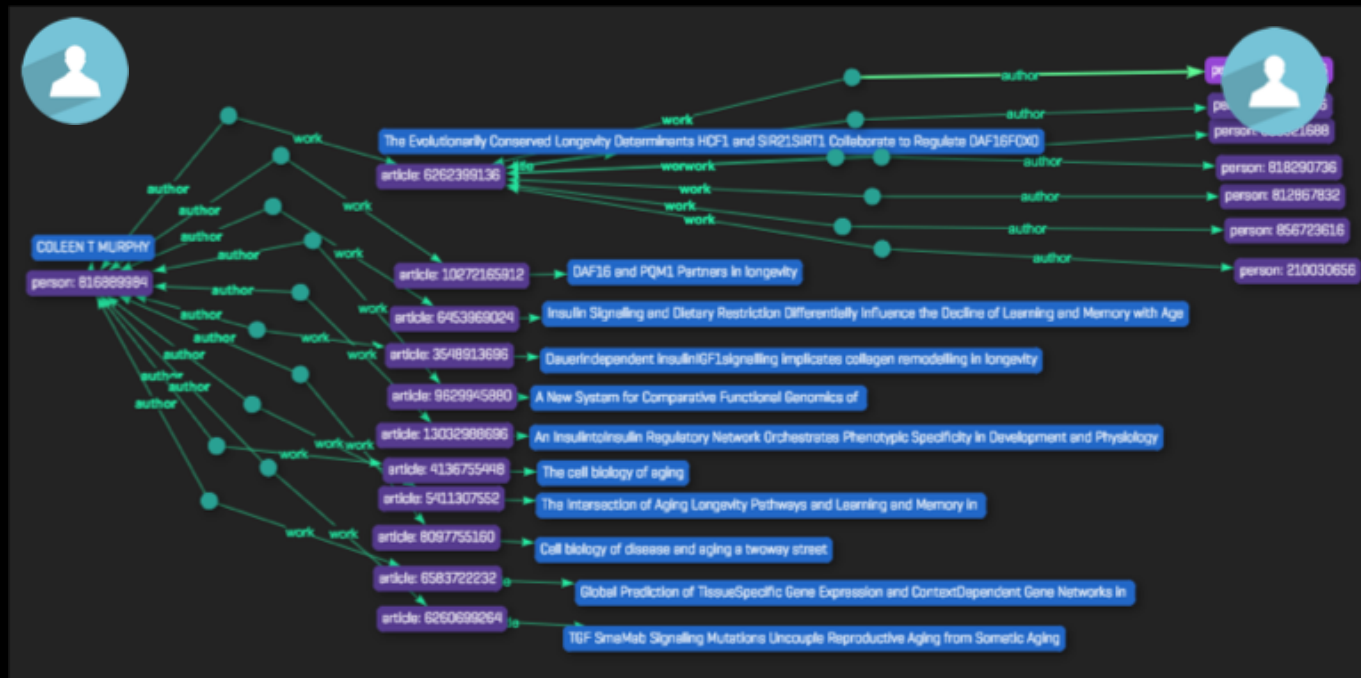


VARIANCE is lower in New York



Can we identify the best scientists by paper count? (most prolific publishers)

Learning about Scientists with a Knowledge Graph



Start with Pubmed data in GRAKN.AI



Title words give an idea of expertise



New York

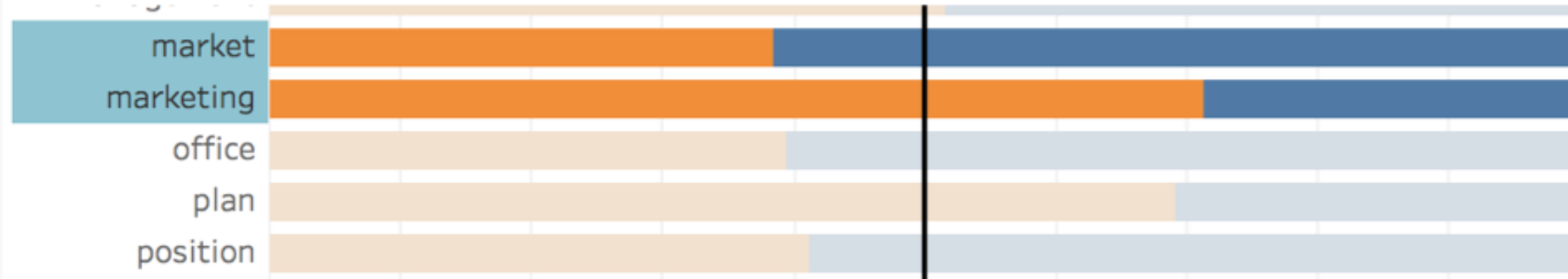


Sheet 1

Region

- New York
- Silicon Valley

Skill

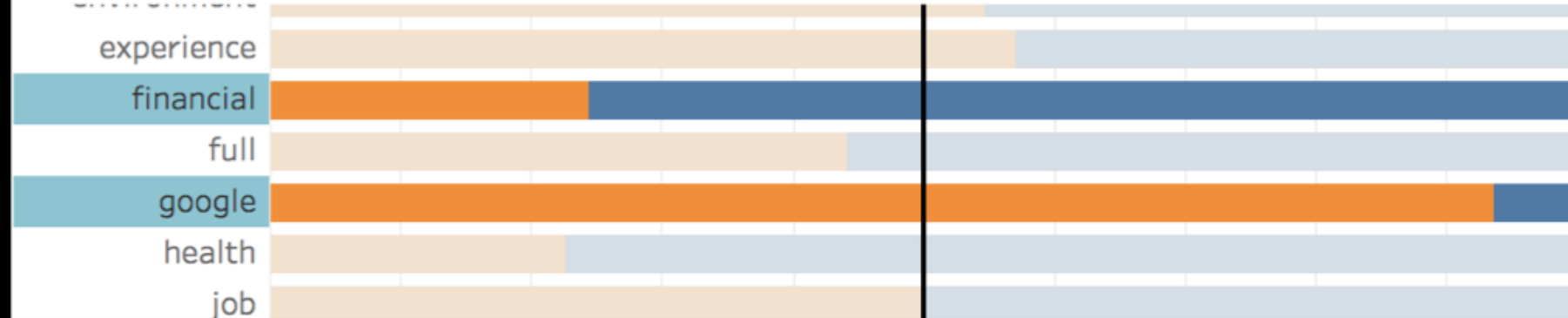


Sheet 1

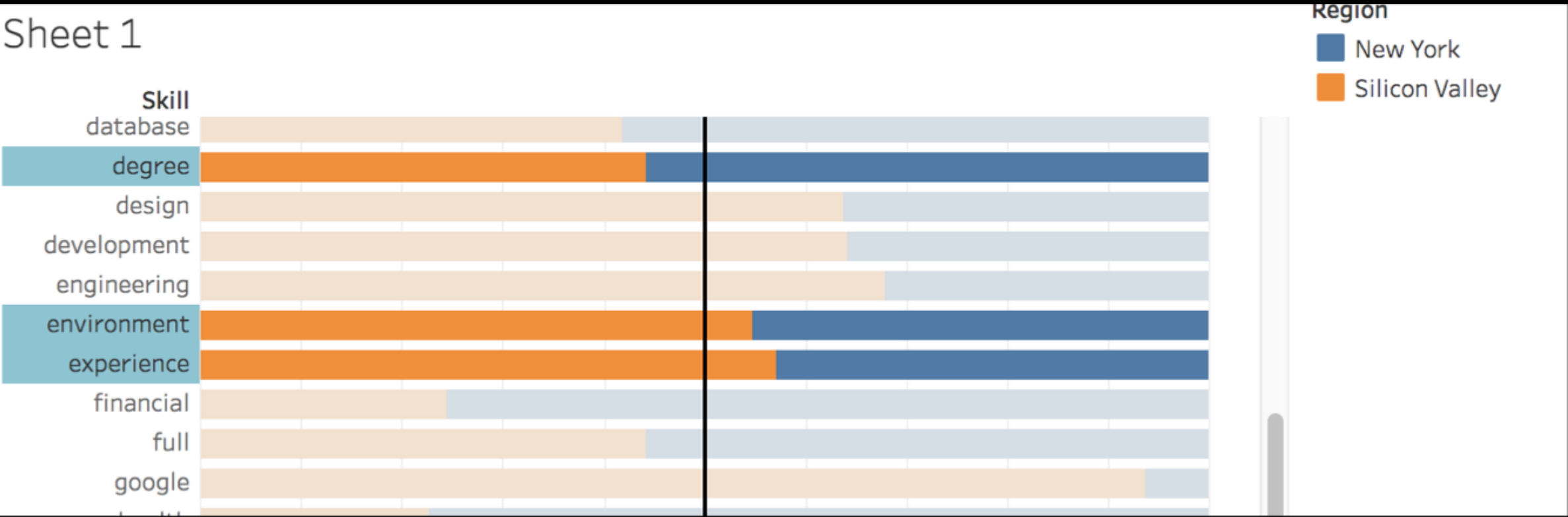
Region

- New York
- Silicon Valley

Skill



Sheet 1



Live demo of pigeon API

(compare “Data Science” related)



So Who wins?

You Decide...



Thank You!

Denis Vrdoljak, MIDS
Managing Director, BDSG
Data Science Instructor, UCB
denis@bds.group

Gunnar Kleemann, PhD, MIDS
Senior Data Scientist, BDSG
Data Science Instructor, UCB
gunnar@bds.group



BACKUP SLIDES



Tableau Link

https://public.tableau.com/profile/denis.vrdoljak#!/vizhome/SVvsNY_JobSkillsPercentCompare_FULL/Sheet1?publish=yes

